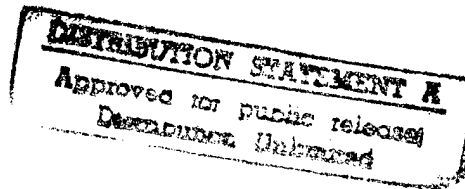# VOLUME III
SYSTEMS PHASE

CHAPTER 2A

SATELLITE NAVIGATION USING
THE GLOBAL POSITIONING SYSTEM

19970117 022

JANUARY 1996

USAF TEST PILOT SCHOOL
EDWARDS AFB, CALIFORNIA

DTIC QUALITY INSPECTED 1

SATELLITE NAVIGATION USING THE GLOBAL POSITIONING SYSTEM

Textbook
Condensed for TPS

Author: Bob Riggins
AFIT/ENG

Designed for 4 hours

1 *All About This Course*

1.1 *Description:*

These notes are designed to accompany a 4 to 6 hour course. They provide a theoretical and practical foundation for understanding the Global Positioning System (GPS). Emphasis is on the use of GPS for determining navigational information such as user position and velocity relative to the local navigation frame of reference (latitude, longitude, altitude, and their time derivatives). Topics include history and motivation for GPS, basic properties of GPS, navigation solution theory, signal structure, code generation, code correlation, receiver design, ranging errors, geometrical errors, differential GPS, relative GPS, and carrier-phase GPS. By the conclusion of this course, the student will be able to write simple positioning algorithms given GPS pseudoranges. Also, the student will become well versed in the theoretical aspects of GPS, and so will be able to read and criticize current GPS research.

Background Reading:

A multitude of GPS books and articles are flooding the market now. A good source for getting a list of all current GPS books is Navtech Seminars and GPS Supplies. Below I list the articles and books I used to write these notes.

1. "NAVSTAR GPS Offers Unpresented Navigation Accuracy" by Denaro   (Hand-out)
2. "Overview" by Parkinson   (Red Book, Volume 1)
3. "Principle of Operation of NAVSTAR and System Characteristics" by Milliken and Zoller (Red Book, Volume 1)
4. "Control Segment and User Performance" by Russel and Schaibly (Red Book, Vol. 1)
5. "NAVSTAR/GPS 18 Satellite Constellations" by Jorgensen (Red Book, Volume 2)
6. "GPS Navigation Using 3 Satellites and a Precise Clock" by Sturza (Red Book, Vol. 2)
7. "Signal Structure and Performance Characteristics" by Spilker (Red Book, Volume 1)
8. "The GPS Navigation Message" by Russell et al (Red Book, Volume 1)
9. "User Equipment Error Models" by Martin (Red Book, Volume 1)
10. "Differential Operation of Navstar GPS" by Kalafus (Red Book, Volume 2)
11. "GPS Navigation: Combining Pseudorange with Continuous Carrier Phase Using a Kalman Filter" by Hwang and Brown   (Red Book, Volume 4)
12. "Guide to GPS Positioning" by David Wells.
13. "GPS: Theory and Practice" by Hofmann-Wellenhof, Lichtenegger, and Collins.
14. "The Navstar Global Positioning System" by Tom Logsdon

Topics to be Covered:

The major topics covered in this text are organized into the following sections:

About this text

The writing of a GPS text evolved from about 5 years of teaching GPS to Air Force Institute of Technology graduate students in the navigation program and to Wright State university graduate/undergraduate students in Electrical Engineering. Consequently, as a teaching tool, that text had been tried, altered, and proven many times over. That text was designed for a 40 hour senior undergraduate/graduate course. These notes, designed for a 4 to 6 hour course, are a condensed version of that text. Advanced topics, exercises, experiments, and some examples are omitted from the original text to form these notes. For more details, you must consult the original text.

2. Motivation and History of GPS

2.1 Motivation

All right, so motivate me to get involved in GPS

  GPS is a space-based radio navigation system giving earth-bound and near-earth users position, velocity, and time estimates. The estimate accuracies vary depending on which technique is employed. For example, using measurements called "pseudoranges" results in a position estimation error of about 100 meters standard deviation. Various techniques are available to improve performance considerably such as Differential GPS, Relative GPS, and Carrier-Phase GPS. For example, exploiting carrier-phase will result in position estimation errors down in the centimeter range.

Other navigation techniques

  Before getting into the nuts and bolts of GPS, we will first examine what other navigation techniques are available. Recognizing the shortcomings of available navigation systems will enable us to appreciate the improvements offered by GPS.

  Many navigation techniques exist. Major categories are as follows:

1. Deadreckoning
2. Piloting
3. Celestial
4. Inertial navigation
5. Radionavigation

Deadreckoning integrates acceleration twice or velocity once to get position. For example, if you drive in one direction at 30 miles per hour, then you've traveled about 0.5 miles down the road in one minute. This method was originally called "deduced reckoning"; however, after many pilots died while attempting to cross the Atlantic back in Lindberg's days using this type of navigation, the "ded" became "dead". As recently as the Vietnam conflict some fighters still used bulky deadreckoning instruments with mechanical servos and gears in place of INSs and electronic computers! Piloting is the fixing of one's position by flying over familiar or marked landmarks. Flying over a town watertower and reading the name of the town is one example of piloting. Piloting may not be the method of choice while flying over the middle of the ocean. Celestial navigation exploits the rigidity of the earth's spin axis in space and knowledge of local time. Latitude is easy to get from looking at the stars (for example, the elevation of Polaris gives latitude directly). Knowledge of time will also give longitude, but small uncertainties in time mean large uncertainties in longitude (.025 mile error for every one second error in the clock--now you know why the ancient navigators worried so much over accurate clocks). Inertial navigation is actually a sophisticated version of deadreckoning;

however, those who use inertial navigation techniques usually do not die. In inertial navigation, we integrate acceleration and/or velocity measurements, maintain knowledge of the instrument frame with respect to a navigation frame, and get user position. For radionavigation methods, we generally exploit the travel time of electromagnetic radiation. Most navigation equipment uses the radionavigation method, including GPS.

The purpose of these notes is not to cover details of other navigation techniques, so I will simply tabulate some of the most important ones (in my opinion) and leave the details to the references. The following table illustrates the quantities that are measured and the problems associated with the listed navigation system. This table will hopefully set the stage for understanding the advantages of GPS. In the table, the term LOS stands for "Line Of Sight" and GDOP stands for "Geometrical Dilution of Precision".
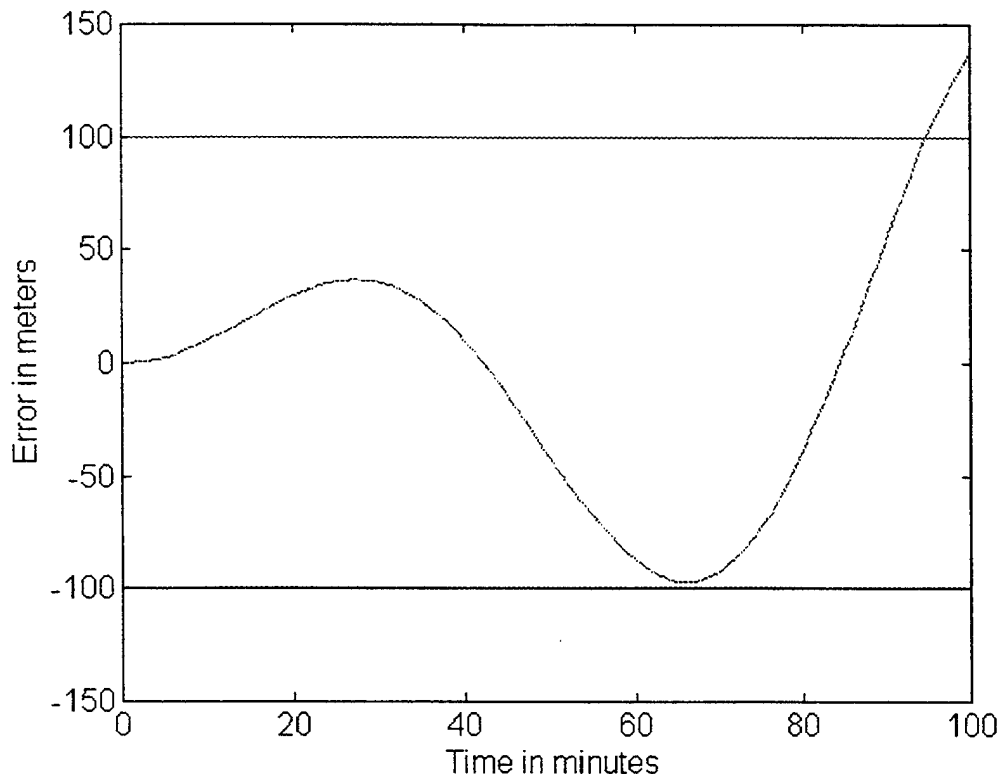
SYSTEM	MEASURED QUANTITY	PROBLEMS
INS	Position, velocity, attitude	Drifts, long-term errors
Loran	Position	Small area, no altitude, poor GDOP, jammable
VOR/DME	Position	LOS, low accuracy
Omega	Position	Foreign control, low accuracy
TACAN	Position	LOS, low accuracy
TAN	Position	High storage, unobservability
JTIDS	Position	Small area coverage
Flyover	Position	Data base, low accuracy at high altitudes
Doppler radar	Velocity	Low accuracy over water and high dynamics

Next I'd like to choose a few of the navigation methods above and analyze their capabilities with respect to GPS. Clearly, I cannot do this for every available navigation method or you'll spend the next 3 months of your life doing this. I choose to compare GPS to Inertial Navigation Systems (INSs) and to VOR/DME. If you're interested you can do similar analyses for the other navigation systems.

What's wrong with INS only?

Why even worry about other navigation systems when we can use an INS? We know a typical INS outputs nearly continuous (typically 100 Hz) position and velocity data, whereas most other navigation systems output at a much slower rate. Accelerometers and gyros measure high dynamics with ease. Furthermore, with the coming of age of optical sensors and micromechnical sensors, the cost and reliability of INSs gets better and better. But the

problem is an INS drifts over time, mainly due to gyro drifts and accelerometer biases. These drifts can be large, depending on the quality of the INS. Typically, we measure accuracy in how far the navigation solution drifts off in 1 hour (ranges from 0.1 to 10 miles per hour drifts). The Litton LN-94 INS we have in the AFIT navigation laboratory drifts at 0.8 miles per hour maximum (1 sigma), and the Rockwell Digital Quartz Inertial Measurement Unit (IMU) has a specification on the maximum drift of 10 miles per hour. The following figure shows a typical drift.

Fortunately, INSs and other navigation systems such as GPS complement each other very nicely. Navigation systems other than INSs usually have very poor performance during high dynamic maneuvers, but are stable on the long run, *i.e.*, they do not drift. INSs are good at high dynamic maneuvers, but are not stable on the long run. So, INSs are at their best when integrated with other navigation systems.

What's wrong with ground-based radionavigation?

You may ask, "What's wrong with ground-based radionavigation? I've been using it for 50 years, so why should I change now? Some new general must have come to power and desired to be remembered for something so he decided to strip away ground-based radionavigation and blast it into space. Humbug." Well, calm down and let me explain.

Basically, the designer of a ground-based radionavigation system has two choices. One, he can use low frequencies and bounce around the world, thereby giving himself worldwide coverage. However, the wavelengths of such systems measure in miles rather than in inches, and atmospheric errors play havoc with such systems, so errors also measure in miles! (A good example is Omega).
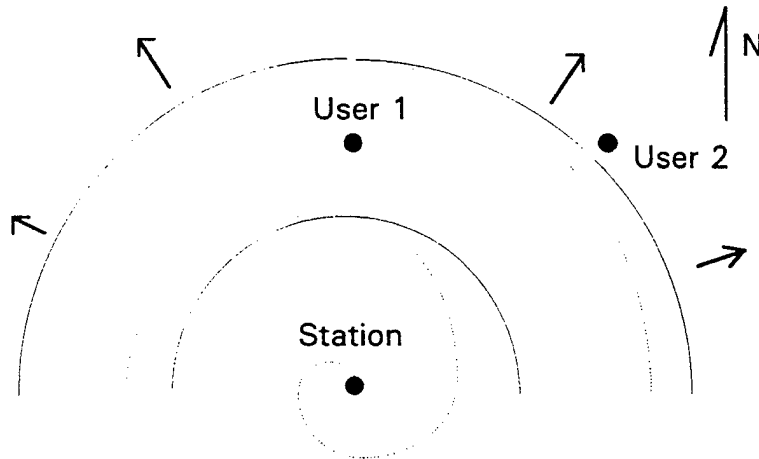
The other choice the designer of a ground-based radionavigation system has is that he can use high frequencies and be limited to line-of-sight. This of course limits the area covered. A complete worldwide system would have umpteen million stations everywhere. Boy, I'd hate to be the one responsible for putting a station at Mariana's Trench in the Pacific Ocean!

At last count, there were over 100 different types of radionavigation techniques available. To study each and compare each technique to GPS would be light years beyond these notes. Instead, I will pick on one popular method, TACAN/VOR/DME, which gives 2-dimensional position with a single measurement and 3-dimensional position when combined with an altimeter. If you desire to study other methods you are stuck with the references.

Shortcomings of TACAN or VOR/DME

The military TACAN (Tactical Air Navigation) and the civilian counterpart VOR/DME (VHF Omnidirectional Range/Distance Measurement Equipment) are well-used and well-known ground-based radionavigation approaches. Both use a "lighthouse" beam and a blinking signal to give range and direction to a known location. The navigator easily finds range by exploiting the speed of electromagnetic radiation as I described before. The sweeping lighthouse beam and blinking signal yield direction as follows:

1. The lighthouse beams sweeps a full circle 30 times a second, and the blinking signal blinks 30 times a second. When the lighthouse signal is pointing due north, the blinking signal is "blinked" on.

2. If the user is due north of the station as User 1 is in the picture below, then the user will see both the lighthouse and the blinking signals simultaneously.
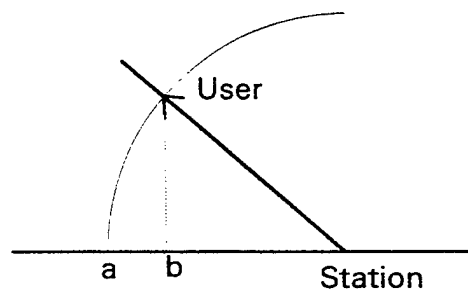
User 1

User 2

N

Station

3. However, if the user is at a position other than due north of the station like User 2 in the picture above, then the user will see the blinking signal before or after the sweeping signal. The time difference between these signals is given by

$$time\ delay = \frac{1}{30} * \frac{\theta}{360}$$

where θ is the angle from north in degrees. Note this method gives position of the user with one VOR/DME measurement in a two dimensional scenario. The problems with VOR/DME include the following (these problems apply also to TACAN except where noted):

1. The navigator needs external altimeter information to get reasonable latitude and longitude data as well as altitude data. The picture below illustrates this problem.

User

a   b   Station

The actual horizontal position is "b" but the indicated horizontal position is "a".

2. A "cone of silence" exists above the station. The lighthouse beam is not omnidirectional. You can visualize this cone by sweeping the slanted line shown in the picture above around the station.

3. VOR/DME requires an active transmitter on the ground for the VOR part and an active transmitter on the aircraft for the DME part.

4. Range is limited to a maximum of 200 miles for high altitude aircraft.

5. The LOS must always be maintained. Mountains can get in the way.

6. For worldwide coverage, stations must be built everywhere, even in enemy territory.

7. The VHF band of frequencies used by VOR can have severe interference problems (TACAN uses much higher frequencies so it does not have as much interference).

8. Large errors exist for large distances. The angle error is about 3 degrees.

Although VOR/DME (or TACAN) is a great system judging by its popularity during the previous several decades, you can see there is much room for improvement.

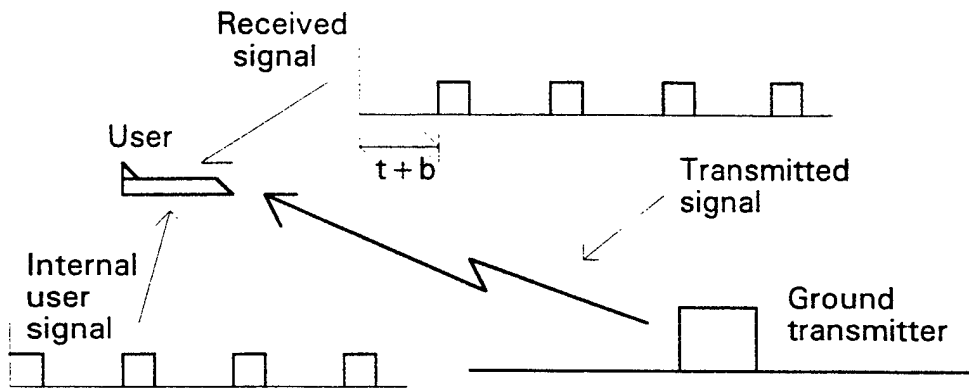*The usefulness of a satellite navigation system*

So how does satellite navigation beat these other techniques? The usefulness of any satellite navigation system in providing positioning information is illustrated by this list:

1. Improved navigation accuracy due to the ability to use high frequencies but not be limited in coverage area. Limitations of ground stations include LOS for high frequencies and large wavelengths for low frequencies.

2. Worldwide coverage is made possible because any potential navigation satellite is at least 100 miles above the surface of the earth.

3. Inexpensive receivers are possible for the user on the ground. Of course somebody has to foot the bill for the satellites.

4. A reduction of the proliferation of other navigation systems is possible because of such overwhelming advantages of the satellite approach.

An accurate satellite-based navigation system would have worldwide coverage (including service for satellites in lower earth orbits) and increased accuracies, and would reduce the need for all the other myriad navigation aids. Moreover, if satellites transmit to ground receivers so that users do not have to transmit, then users can be passive and not traceable by the enemy. Also, the satellites can transmit their signals continuously to make navigation possible any time of the day anywhere on the earth.

Passive radionavigation:

GPS uses passive radionavigation. In passive radionavigation, the user only receives. Typically, the ground transmitter sends a known periodic signal on a continuous omnidirectional basis. The user receives the signal and compares the delay of the signal to its own internally-generated signal. If the user clock and ground transmitter clock are not perfectly synchronized, then the delay term will have an unknown user clock bias, b. In the following picture, the ground transmitter transmits signals continuously in all desired directions, and the aircraft receives those signals continuously.

The range from the ground transmitter to the user is

$$range = c(T - b)$$

where $T = t + b$ is the measured delay of the signal and t is the unknown true delay. Obviously, you cannot solve for range with just this equation since b is unknown. This type of navigation essentially increases the number of unknowns by one. But with 4 measurements like this, we have 4 unknowns and 4 equations.

The advantages of passive radionavigation are that the user equipment is relatively simple and the user does not transmit. Also, the user can compute accurate time. The disadvantages are that the user clock has an unknown bias (thereby increasing uncertainty and requiring more measurements) and the ground transmitter is vulnerable.
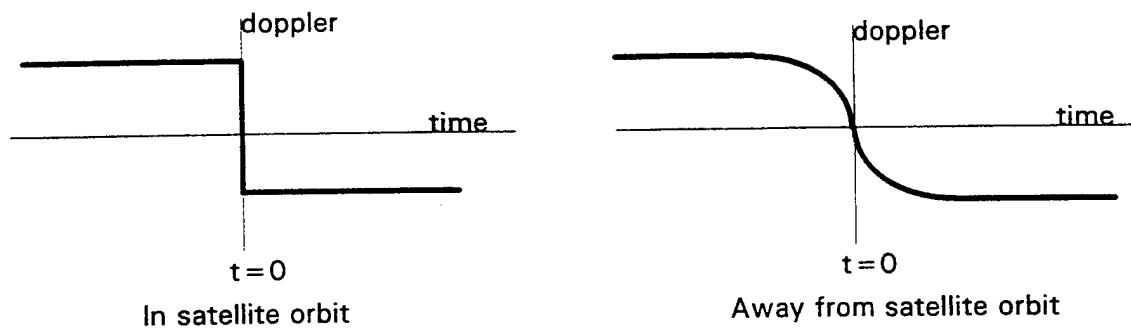
2.2 History of GPS

The other references I listed at the beginning explain the history of GPS very well. Whereas the history of GPS may be informative and almost exciting, history is not the primary subject of these notes. However, just for the sake of completeness, I will include a brief outline of the more important and interesting GPS historical facts.

Naval National Satellite System (NNSS) TRANSIT

1. The idea started from observing signals from Sputnik in the 1950's. The entire orbit of Sputnik could be determined by observing the doppler as the Soviet spacecraft passed overhead. Scientists then got the bright idea of inverting this technique to produce a satellite navigation system based on doppler.
2. Initial TRANSIT satellites lasted only 6 to 8 months. Nonetheless, ships are still using TRANSIT today (1995).
3. TRANSIT satellites are in 580 mile high polar orbits. This means the periods are about 1.5 hours and each pass lasts from 10 to 15 minutes (or less).

4. TRANSIT was originally developed for the Polaris submarine fleet. In between dives, the submarines could obtain position fixes while at the surface.

5. Global intermittent coverage is possible with TRANSIT with 5 or 6 satellites.

6. The principle of operation of TRANSIT is as follows: The satellites transmit a continuous tone. Receivers in turn pick up the signal and determine position from doppler by measuring the signal over a time interval. The shape of the doppler curve gives the distance from the satellite orbit:

In satellite orbit

Away from satellite orbit

7. Gravity gradient stabilization is the chosen method to maintain TRANSIT pointing towards the earth. They do this by using a 50-foot telescoping beam.

8. The disadvantages of TRANSIT are:

>   Only latitude and longitude fixes (no vertical)
>   Available once per hour
>   Takes 10 to 15 minutes to get fix
>   Need for independent speed and altitude
>   Inaccurate near the poles

US Navy TIMATION

1. The Navy wanted to improve on TRANSIT .

2. The name comes from TIMe and navigATION.

3. The biggest motivation was that the Naval Research Laboratory (NRL) was also interested in time as well as position. Recall time is readily obtained with a passive approach.

4. TIMATION satellites were set at middle altitude orbits (7500 nautical miles).

US Air Force 621B

1. The Air Force did this program the same time as Navy did TIMATION--what competition! Of course we all know the AF program was better...but the Navy was better at descriptive and imaginative names.

2.   The 621 satellites are synchronous at 19000 nautical miles altitude. Note, they are not geosynchronous so their ground traces are in the shape of figure eights centered on the equator.

3. This approach was much better for high dynamics, which is what the Air Force would naturally be interested in!
4. They were the first to use the pseudo-random noise code. GPS also uses this as we will see.
5. One big problem: they required ground transmitters. Which leads to the next thing:
6. This 621B approach was very vulnerable to attack!

GPS

        The Air Force then picked the best characteristics of all these previous programs to build the NAVSTAR/GPS. GPS has the following characteristics:

        1. Global coverage (with at least 24 satellites)
        2. High accuracy 3-D position capability
        3. Time information
        4. Continuous availability
        5. Passive service (users do not transmit)
        6. Unlimited number of users
        7. Resistance to interference and jamming
        8. Altitude of 11,000 nautical miles
        9. All weather capability

Users do not have to transmit anything, and the satellite signals are resistant to jamming due to the spread spectrum property of the pseudo-random noise code.

        I hope this brief discussion on GPS history is adequate to meet your needs. If not, too bad. Check the references for more details. Next, we'll discuss the basic properties of GPS.
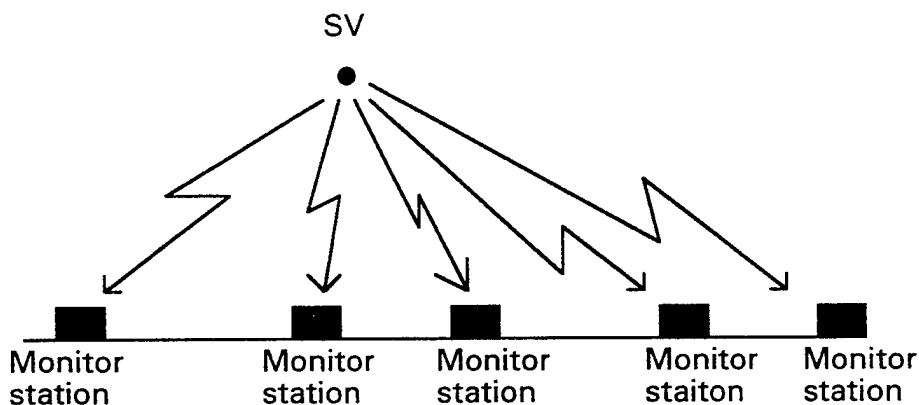
3. *Basic Properties of GPS*

The basic properties of GPS are best described by partitioning GPS into 3 fundamental segments:

1. Control segment
2. Space segment
3. User segment

The control segment, including stations around the world, controls the satellites and calculates errors in satellite orbits and clocks. The space segment includes the satellites themselves as well as the signals involved, and the user segment includes the receivers and users like you and me.

3.1 Control Segment

The purpose of the control segment is to continuously monitor the satellite positions and clocks and determine correction factors. Control inverts the ranging process that we will use in Section 5. Five monitoring stations spread over the globe receive the same signals you and I do, and find pseudoranges just as you and I. The difference is that we know exactly where these monitoring stations are, and by ranging to one satellite with all these stations, we can back out *satellite* position and time. In order to avoid poor geometry, the monitor stations must be spread over the globe (an alternate method proposed is for each of the satellites to monitor each other). This inverted ranging process concept is depicted below:

SV

| Monitor station | Monitor station | Monitor station | Monitor staiton | Monitor station |

Five monitor stations are located at

1. Hawaii
2. Diego Garcia
3. Ascension Island
4. Kwajalein

  5. Colorado Springs

As you can see, the five monitoring stations are spread over the entire world. These monitoring stations collect pseudoranges and send them to the master control station at the Air Force's Consolidated Space Operations Center located near Colorado Springs (Falcon Air Force Base). The master control center uses a combination of Kalman filtering and least squares to compute the orbits and timing of each GPS satellite. Now, master control can compute the difference between where the satellites think they are and where master control says they are. These corrections are then transmitted to upload stations at

  1. Ascension
  2. Diego Garcia
  3. Kwajalein

Note that these upload stations are also monitoring stations. The upload stations periodically (daily for now, but that period may change) send these corrections up to the satellites on S-band using 16- foot ground antennas. The satellites in turn divide the corrections into hourly corrections and transmit them to users in the data message which I'll describe shortly. Now, the users have up-to-date information on the satellite whereabouts and satellite clocks.

3.2 Space Segment

  GPS satellite construction comes in blocks of all sizes. For instance, there are Block 1, Block 2, Block 2r, and Block 2f GPS satellites. Block 1 satellites were used in the initial test program. The following list highlights the important information on the Block 1 satellites:

  1. Total of 12 satellites
  2. Tests done at the Army Yuma Proving Grounds in Arizona
  3. NAVSTAR 7 blew up when its Atlas booster exploded over VAFB
  4. NAVSTAR 12 never went into space
  5. As of 1992, 5 Block 1 satellites were still operational.
    As of 1993, 4 were in use.
    In 1994, we were still seeing 3.
    In 1995, I stopped seeing any more Block 1 satellites on my receivers.
  6. Rockwell International was the builder
  7. Each satellite weighed 960 pounds
  8. Orbits were inclined 63 degrees to equatorial plane\3 orbital planes

It is important to note that the reliability of the Block 1 GPS satellites was much higher than they were designed for.

  Block 2 satellites are the current main force. The following list highlights information on the Block 2 satellites:

1. The Block 2 satellites had these physical parts:

   Thermal control
   Solar panels
   Navigation payload
   Orbit insertion
   Global burst detector
   L-band transmitter antenna
   S-band TTM receiver antenna

2. Final constellation is 21+3 spares
3. 6-11 Satellites always in view ≥ 5 degrees elevation
4. Orbital period is about 12 hours. Consequently, we have
       (a) Gradual degradation in case some are shot down
       (b) Satellites in view for about 3 or 4 hours
5. Altitude about 10,900 nmi
6. Launched by the shuttle and expendable launch vehicles (elv)
       (Atlas/Delta)
7. The builder is Rockwell International.
8. Each satellite weighs about 2000 pounds
9. Orbits have 55-degree inclinations / total of 6 orbital planes / each plane
       60 degrees apart

Block 2R satellites are the replenishment satellites:

1. Each satellite will weigh 2300 pounds
2. Satellites will be built by General Electric
3. Launch program will begin soon

Launch schedule for Block 2

To give you an idea of the launch frequency, I've included the following tidbits:

11 launched as of July 1991
16 total (both blocks) up July 1992
Other various launches between then and now
24 satellites up now
GPS World is a good source for keeping up with this

Outage possibilities

Outages occur when satellite coverage is insufficient to do navigation. During the 1991 and 1992 time frames, not having enough satellites to do 3 dimensional navigation was a real problem. Now of course we always have enough satellites in the sky to do navigation;

however, due to shading and other causes for loss of lock, we still have that possibility of an outage. Two major types of outages are:

1. Poor geometry
2. Less than 4 satellites--need at least 4 for navigation solution

We will cover geometry in detail in another section. For the second type of outage listed above, you're on your own--maybe you should get out from under those trees, get away from the skyscrapers, or just buy a better receiver.

Frequencies

The carrier is transmitted at 2 L-band frequencies in the lower microwave section of the electromagnetic spectrum. These frequencies are

L1   1575 MHz
L2   1227 MHz

Having the signal available on two frequencies allow for ionosphere correction.

Codes

The carrier is modulated with one or two pseudo-random noise codes that spread the signal in frequency (spread spectrum). The use of such codes allows us to pick up GPS signals even though the signals can be as much as 30db below the noise power level!

The carrier signals are modulated with 2 codes:

1. P-code (P for precise) or Y-code (encrypted version of P-code)
2. C/A-code (Clear or Coarse/Acquisition)

L1 has both codes and L2 only has P-code (note, in the beginning L2 carried both P and C/A codes). The Y-code is to exclude everyone except authorized users like the US military from using the more highly accurate P-code. This encryption is sometimes referred to as "anti-spoofing" or AS. So, GPS provides 2 navigation services:

1. Standard Positioning Service (SPS). This is C/A code only.
2. Precise Positioning Service (PPS). This is use of P-code (or Y-code when encryption is turned on).

Since decryptors are limited to authorized users, most users can only use SPS.

Codes are important in GPS since each satellite is distinguishable from the others by its code. In the following two lists I include some brief code facts to whet your appetite.

P-code

1. Pseudo-random noise (PRN) code of 7 days period
2. Initiated once a week
3. Hard to lock on
4. Chip rate is 10 times faster than C/A code, so accuracies are higher
5. "HOW" (hand-over-word) transmitted to users every 6 seconds to ease acquisition
       (HOW indicates satellite time of transmission)
6. P-code allows use of L2 which in turn allows for ionosphere correction

C/A-code

1. PRN of .001 seconds period
2. Easy to lock on
3. Get HOW from data
4. Not as accurate as P-code
5. Ambiguity possible (could lock on to the wrong code cycle)

GPS time

It is important that all the satellite clocks are synchronous, or at least we should know how far off each satellite clock is relative to the others. Each satellite has 2 rubidium and 2 cesium atomic clocks. The navigation solution depends on the precise synchronization of all Satellite Vehicle (SV) clocks. As you will see in a later section, we need to know when the satellite signal is transmitted in order to calculate the navigation solution. Of course, we're not super-humans, and so we cannot know precisely when the signals are transmitted, but we have to be close. How close? One nanosecond of timing error produces:

$$range = speed \times time = (3 \times 10^8 \, m \, / \, sec) \times (1 \times 10^{-9} \, sec) = .3m \approx 1 \; foot$$

So, timing is important.

Some other satellite clock tidbits of information you may or may not want to know are:

1. Control station takes an average of a number of cesium clocks
2. Monitor stations take range and time measurements from satellites
3. Control updates the satellite clocks with its own corrections
4. Clocks calibrated by control daily
5. These corrections are sent to satellites and satellites send to users
6. Receivers then extract satellite clock biases
Note, you'll be doing this--without these corrections, your navigation solution could be as far off as several hundred kilometers!

7. GPS time≠UTC time (Universal Coordinated Time)--Hey, it's not my fault the letters are switched! There are no leap seconds in GPS time.

8. Fundamental clock frequency is 10.22999999545 MHz at satellite. This looks like 10.23 MHz to us because of relativistic effects due to satellite velocity (about 2.5 miles per second), Earth's gravity, and Earth's rotation. All other frequencies are derived from the fundamental frequency, F:

P-code chipping rate.........................10.23 MHz (F)

C/A code chipping rate.....................1.023 MHz (F/10)

L1 carrier.......................................154x10.23 = 1575.42 MHz
    (Fx154)

L2 carrier.......................................120x10.23 = 1227.6 MHz
    (Fx120)

Chip width is defined as the code bit width.

More on degradation: What's this AS/SA flak?

AS is anti-spoofing = encryption of P-code (creates Y-code). With a C/A code receiver, you won't have to worry about Y-codes and such.

SA is Selective Availability = intentional degradation of satellite signal so commercial users and unfriendlies cannot enjoy total accuracies of GPS. A dither signal is applied to the satellite ephemeris data and/or the satellite clock. SA only affects C/A code--typical error due to SA is 100 feet. (A few years ago during the Persian Gulf conflict, they were turning SA off and on, and it was easy for me to tell the difference in my navigation solutions!) Differential and relative GPS largely remove SA--don't worry, you'll have the pleasure of doing all of this.

Data Message

Both the L1 and L2 signals are modulated with at least one code and the data message. For example, the L1 signal has

Carrier + C/A code + P-code + Data message

The data message rate is 50 bits per second and includes:

1. Satellite Vehicle (SV) status (Is the satellite healthy?)
2. Hand-Over Word (HOW)
3. Clock corrections that Master Control has updated to the satellite
4. Ephemeris (satellite position and velocity information)
5. Other corrections such as atmospheric
6. Almanac data for finding other satellites to make acquisition simpler

This data message is on both L1 and L2. Receivers need to extract information from the data message in order to write your navigation algorithms. More on this later.

3.3 User Segment

This includes you, me, Aunt Molly and the receivers. You know about you, I know about me, and who knows about Aunt Molly, so let's talk receivers. Basically, I like to break the receivers into 4 parts:

1. The receiver antenna and electronics
2. The code and carrier tracking loops, and code acquisition circuitry
3. Navigation computer
4. Control display unit

In this course, we'll discuss all four, but the bulk of our work is concentrated on navigation computations.

The antennae for GPS are designed to receive right-hand circular-polarized signals. These antennae are usually omni-directional and come in many shapes. We have as many antenna shapes as we do GPS receivers. The L1 and L2 signals are normally routed through a preamp that boosts their power level and then sent through cable (such as co-axial) to the receiver. Typically, DC power to the preamp is supplied to the antenna on the same line containing the signal. For instance, the signal may ride on a 12 volt DC signal in the coaxial cable (I found this out the hard way).

The code, tracking, and acquisition electronics are covered in other references, but for now, get this burned into your brain. The purpose of the code loop is to provide the navigation computer the pseudorange measurements (pseudorange explained shortly), and the purpose of the carrier loop is to provide doppler and the data message as well as carrier-phase information.

Specially equipped receivers also have the capability to measure carrier-phase in the carrier loop. You'll see that this is a neat capability to have, since we can knock our errors down by an order of magnitude with carrier-phase.

The navigation computer takes all the raw measurements, like pseudo-range and doppler and carrier-phase, and computes latitude, longitude, height, velocity, etc. Even attitude can be computed with multiple antennas.

The control display unit gives current position and velocity information to the user. Also, it has other functions like the display of time and waypoint navigation instructions.

The experiments and exercises we do at AFIT are developed using Navsimm's XR series of GPS receivers. If you happen to use the older XR4 receiver, you'll notice the XR4 is

much slower than the XR5 during acquisition. This is because the XR4 is a multiplexing receiver, having only two channels. It multiplexes in time by dwelling on each satellite briefly, and then moves on to the next satellite. It takes a long time to initially find good satellites in this sequential manner. The XR5 on the other hand is a continuous tracking device, with 6 parallel channels for the XR5-6 and 12 parallel channels for the XR5-12. This characteristic enables the XR5 to dedicate each channel to one satellite. No switching is necessary and consequently the acquisition time is much faster.

Before going into building our navigation equations, this is a good time to introduce the errors we'll have to face. Nothing's perfect, not even GPS! Errors basically come from two categories in GPS: measurement errors and geometry. User clock error is the biggest error source. In fact, it is so big that in our navigation algorithms we'll treat time as one of our variables, just like position. Other typical 1-sigma measurement errors for P-code are:

SV Clock and ephemeris errors	1.5 meters
SV equipment delay	1 meter
Atmospheric delays	2.4 - 5.2 meters
Multipath	1.2 - 2.7 meters
Receiver noise and vehicle dynamics	1.5 meters

The Root-Sum-Square (RSS) error here is 3.6 - 6.3 meters. Geometry effects typically multiply this number by 3 (we'll discuss that more later). So the total 1-sigma error is typically 10.8 - 18.9 meters assuming the user clock parameter is estimated without error. A final error that I did not mention above is selective availability, which is bigger than all the previous errors combined. One day soon selective availability will be history and I will just delete this sentence from these notes!

Now what exactly is this *pseudorange*? Pseudorange is true range plus all the measurement errors (when code-phase measurements are used). In other words, pseudorange is the receiver *measurement* of the range. By far the biggest factor in the difference between pseudorange and true range is the user clock error. And boy can this difference be big! With the typical receivers we have in the lab, I typically get pseudoranges to be as much as 5% or more off from true ranges.

Of course, other measurements exist besides code-phase pseudorange. In the next section, we'll discuss all available GPS measurements in detail.

4. *Available Measurements*

In this section, you will learn (if the creek don't rise) about all available GPS measurements and how to use them.

On a fundamental level, there are only two available GPS measurements:

1. Code-loop phase
2. Carrier-loop phase

These measurements are physically available on most GPS receivers since each receiver usually has a code-loop and a carrier-loop.

The first fundamental measurement above yields code-phase range, usually called pseudorange. This is the most common and simplest measurement that a GPS receiver uses to compute navigation parameters (position and velocity). Sometimes GPS users (such as us) approximate LOS rate of change by differencing successive pseudorange measurements. We call this "measurement" the pseudorange delta-range, which is not really a measurement but a calculation from the first fundamental measurement.

The second fundamental measurement above yields carrier-phase range. Carrier-phase ranges are much more accurate than pseudoranges. Only more sophisticated receivers have the capability to measure carrier-phase. On the other hand, most receivers measure doppler, the change in carrier-phase from one time to another. Just as with pseudorange delta range, doppler approximates true LOS range rate. The actual doppler measurement uses a cycle counting process divided by a time interval. Since doppler is based on carrier-phase and carrier-phase is much less noisy than code-phase, then doppler approximates range rate better than pseudorange delta-range.

Maybe it would help to sum all this up in a table:

Phase measurement	Range-Rate Approximation
Code-phase range (direct measurement)	Pseudorange delta-range
Carrier-phase range (direct measurement)	Doppler (carrier-phase delta-range or just delta-range)

Almost all receivers have the capability to measure code-phase range and doppler because these are relatively simple for receivers to do. Some more expensive receivers also measure carrier-phase range.

21

Next, I'll walk you through the equations associated with each of these 4 measurements or calculations from measurements.
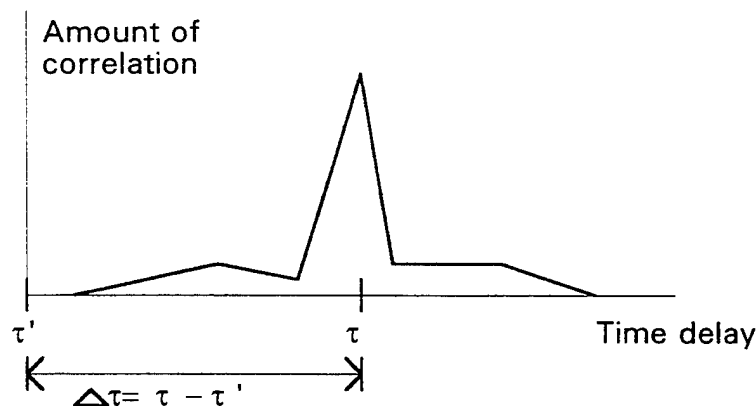
4.1 Code-Phase Measurement

This is the familiar pseudorange, which is the most common GPS measurement. With code-phase alone we can compute our position and velocity. However, it may not be as accurate as if we used all available measurements.

How does a receiver measure pseudorange?

GPS receivers measure the phase shift between a receiver-generated signal and the signal received from the satellite.   The receiver shifts its internally generated code until it maximally correlates with the received signal. Each satellite has its own unique code and at the receiver the code rides about 30 dB below the noise level. Therefore, a receiver can "see" the incoming code only after this shifting and correlating process.   After the code loop shifts the internally generated code and achieves maximum correlation, then the difference between satellite transmission time and receipt time (after the shift) times the speed of light gives us the pseudorange.   Satellite transmission time comes to us in the ephemeris data as the HOW, a 17-bit satellite time indicator using 1.5 second z-counts referenced to the leading edge of the next subframe.

The following picture illustrates the idea of maximum correlation:

In the picture, $\Delta\tau$ represents the amount of shift in time units necessary to achieve maximum correlation. In equation form we have

$$\tau' = t'_r - t_s$$

where t'_r represents received time according to receiver clock before the shift, and t_s is the transmitted time according to the satellite clock. We also have

$$\tau = t_r - t_s$$

The time t_r is the received time according to receiver clock after the shift. Note by subtracting the previous two equations, we can write the shift as

$$\Delta \tau = \tau - \tau' = t_r - t'_r$$

Thus τ represents measured time of transmission of the signal in going from satellite to user. Pseudorange from satellite i results from multiplying the measured time of transmission with the speed of electromagnetic radiation:

$$PR_i = c\tau \qquad (4.1)$$

where $\tau = t_r - t_s$.

The measured time of receipt of the signal equals actual GPS time of receipt plus the user clock error:

$$t_r = GPS\ time\ of\ reception + \delta_{uc}$$

The indicated time of transmission of the signal equals actual GPS time of transmission plus the satellite clock error:

$$t_s = GPS\ time\ of\ transmission + d_{sc}$$

Let:

$$\tau_a \equiv GPS\ time\ of\ transmission$$
$$\tau_b \equiv GPS\ time\ of\ reception$$
$$\delta_A \equiv Atmospheric\ time\ delay$$
$$\delta_{uc} \equiv Offset\ of\ receiver\ clock\ with\ GPS\ time$$
$$d_{sc} \equiv Offset\ of\ satellite\ clock\ with\ GPS\ time$$
$$\delta_{ei} \equiv Ephemeris\ errors\ in\ units\ of\ range$$

The pseudorange (or code-phase) measurement equation is

$$PR_i = c(t_r - t_s) + errors$$
$$= c(\tau_b - \tau_a) + c\delta_{uc} - cd_{sc} + errors \qquad (4.2)$$

where $c(\tau_b - \tau_a)$ would be the true range from user to satellite if the signal traveled through a perfect vacuum with no atmospheric delays. The term $c\delta_{uc} - cd_{sc}$ represents the difference between the user clock offset and the satellite clock offset. Other errors include receiver correlation errors and satellite equipment delay errors.

Next, let's discuss the user clock offset. Two scenarios are possible:

1. The user clock is fast (or advanced) and so $\delta_{uc} > 0$. In this case, the pseudorange is bigger than true range because of the user clock offset, and the offset is called an *advance*.

2. The user clock is slow and so $\delta_{uc} < 0$. In this case, the pseudorange is smaller than true range because of the user clock offset, and the offset is called a *delay*.

Now, concentrate on the satellite clock offset, d_{sc}. You can also make the same statements about the satellite clock offset as we did about the user clock offset above. But besides that, satellite clock delay is actually the sum of two parts:

$$d_{sc} = delt + \delta_{sc} \tag{4.3}$$

The first part, *delt*, is a deterministic part that we can extract from the satellite data stream. The second part, δ_{sc}, is the residual satellite clock offset that we cannot account for.

If the signal traveled through a perfect vacuum with no atmospheric delays, then $c(\tau_b - \tau_a)$ would be the true range from user to satellite. However, because of the presence of the atmosphere, c is not the true speed of the satellite signal. The true range is actually the measured time of transmission *minus* any time delays caused by the atmosphere, all multiplied by the speed of light through a vacuum. Thus, we have

$$c(\tau_b - \tau_a) = true\ range + c\delta_A$$
$$= \left\| \vec{R}_{ti} - \vec{R}_u \right\| + c\delta_A \tag{4.4}$$

where $\vec{R}_{ti}$ is the true satellite position vector referenced to the center of the earth, and $\vec{R}_u$ is the true user position vector referenced to the center of the earth. The symbol $\| \ \|$ represents the Euclidean norm so $\left\| \vec{R}_{ti} - \vec{R}_u \right\|$ is the true range between user and satellite. We must compute $\vec{R}_{ti}$ from ephemeris, and there are always ephemeris errors. In other words, the satellites do not know their positions precisely despite the valiant efforts of GPS's control segment. The best we can do is

$$\left\| \vec{R}_{ti} - \vec{R}_u \right\| = \left\| \vec{R}_i - \vec{R}_u \right\| + \delta_{ei} \tag{4.5}$$

where $\vec{R}_i$ is the satellite position vector calculated from ephemeris data. So, the pseudorange equation becomes

$$PR_i = \left\| \vec{R}_i - \vec{R}_u \right\| + c\delta_{ion} + c\delta_{trop} + \delta_{ei} + c\delta_{uc} - cd_{sc} + other\ errors \tag{4.6}$$

Example 4.1

Suppose a receiver measures range between user and satellite. The measurement of the range, called pseudorange, has the following errors:

$$\delta_{uc} = .002 \text{ sec}$$

$$delt = .001 \text{ sec}$$

$$\delta_{sc} = -3e - 8 \text{ sec}$$

$$\delta_{trop} = 1e - 7 \text{ sec}$$

$$\delta_{ion} = 3e - 7 \text{ sec}$$

$$\delta_{ei} = -1.5 \text{ meters}$$

$$Code\ loop\ error = 3\%\ C/A\ code\ chip\ width$$

What will the pseudorange measurement of the receiver be if the range between true user position and the position of the satellite as indicated by ephemeris data is 27,000,000 meters? If all other errors except code-loop error are somehow eliminated, what is the difference between pseudorange and range?

Solution:

From Equation 4.6 we have

$$PR_i = 27,000,000 + 90 + 30 - 1.5 + 600,000 + 300,000 + 9 + code\ loop\ error$$

$$code\ loop\ error = .03(300) = 9\ meters$$

$$PR_i = 36,000,136.5\ meters$$

♠

4.2 Carrier-Phase Measurement:

Just as we did for code-phase, we can correlate the received carrier to a carrier signal generated inside the receiver. Only now, instead of matching up 300-meter chips from code, we match up 19-centimeter cycles of carrier. Less expensive receivers cannot perform carrier-phase measurements, but good receivers can theoretically measure phase with an accuracy of 1% of cycle, or about 1.9 millimeters! Clearly, there are ambiguity problems here.

Let the indicated times for signal receipt and transmission be:

t_s = Transmitted time according to satellite vehicle clock
t_r = Received time according to receiver clock

The difference between transmitted time and received time is
:

$$\tau = t_r - t_s \tag{4.7}$$

The phase of the internal receiver oscillator at time t will be $\Phi_g(t)$, where the subscript g stands for "generated" signal. The phase of the receiver oscillator at t_r is related to the phase of the oscillator at t_s by the Taylor Series expression:
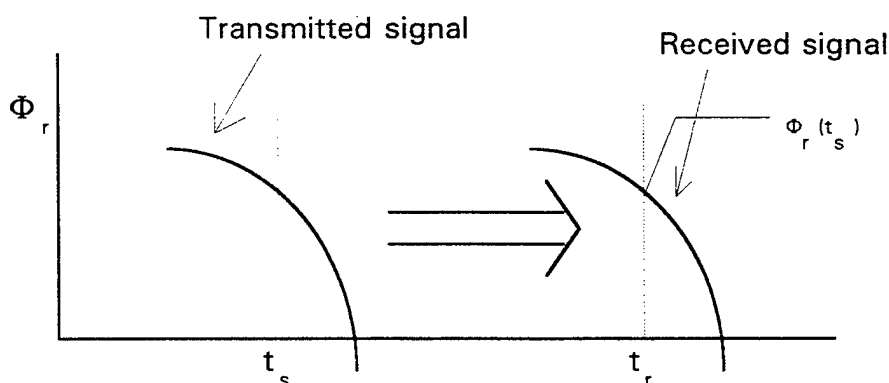
$$\Phi_g(t_r) = \Phi_g(t_s) + \frac{\partial \Phi_g}{\partial t}\bigg|_{t=t_r}(t_r - t_s) + HOT \qquad (4.8)$$

where HOT represents higher order terms in $t_r - t_s$. Notice that frequency of the carrier is the derivative of the phase. For stable oscillators, the frequency remains constant so higher derivatives of phase are zero. Thus, for stable oscillators, we have

$$f = \frac{\partial \Phi_g}{\partial t}\bigg|_{t=t_s} = \frac{\partial \Phi_g}{\partial t}\bigg|_{t=tr}$$
$$\Phi_g(t_r) = \Phi_g(t_s) + f(t_r - t_s) \qquad (4.9)$$
$$\Phi(t_r) = \Phi(t_s) + f(t_r - t_s)$$

where f is the frequency of the oscillator. The third equation simply indicates that this expression is valid for any stable oscillator.

The phase of the received signal transmitted at time t by the satellite will be $\Phi_r(t)$. Note that my notation says *transmitted at time t*, so this phase is received sometime later than t because of the finite speed of the signal over the long distance. Let $\Phi_r(t_s)$ be the phase of the transmitted signal at time t_s. $\Phi_r(t_s)$ is also the phase of the received signal at time t_r. This is what the phase of the signal was when it was transmitted at time $t_s = t_r - \tau$. The picture I am about to draw will clarify this:

Think of this as a surfer riding a wave. As the surfer rides just in front of the crest of the wave he passes a point which we'll call the time (t_s) and place (satellite) of transmission. Later, the surfer is still riding the moving wave just in front of the crest but he passes another point we'll call the time (t_r) and place (receiver) of reception.

Recall, $\Phi_g(t_r)$ is the phase of the signal generated internally in the receiver at time t_r. This is the receiver's attempt to emulate what the phase of the signal at the satellite is at time t_r. If the receiver clock was perfectly synchronous with the satellite's clock, then the phase of the signal generated internally in the receiver would be the same as the phase of the signal being transmitted at the satellite. Of course, the clocks are not synchronous so we will have to account for clock errors just as we did in code-phase.

The phase difference at the receiver at time t_r between the received signal and the signal generated internally in the receiver is given by

$$\Delta\Phi \equiv \Delta\Phi(t_r) = \Phi_r(t_s) - \Phi_g(t_r) \qquad (4.10)$$

where $\Phi_r(t_s)$ is the phase of the received signal at time t_r (or equivalently, the phase of the transmitted signal at time t_s) and $\Phi_g(t_r)$ is the phase of the generated signal at time t_r. The signal generated in the receiver is corrected for doppler through a phase-locked loop so the received and generated signals have the same form, except one is delayed in time with respect to the other. Assuming the two signals are the same signal shifted in time, we will work with one signal instead of two and drop the subscripts. From Equation 4.9 we have

$$\Delta\Phi = -f(t_r - t_s) \qquad (4.11)$$

where we assumed the oscillator is stable.

We also know that the true GPS time of receipt is equal to the true GPS time of transmission plus the time it takes for the signal to travel the true distance between satellite and user. Furthermore we know that the time it takes the signal to travel the true distance is equal to the time it takes the signal to travel that same distance in a vacuum plus path delays such as atmospheric delays. Putting all this mumbo jumbo into an equation so people like you and me can understand, we have

$$t_r - \delta_{uc} = \textit{true GPS time of receipt}$$
$$t_s - d_{sc} = \textit{true GPS time of transmission} \qquad (4.12)$$
$$t_r - \delta_{uc} = t_s - d_{sc} + \frac{\left\|\bar{D}_i\right\| - c\delta_{ion} + c\delta_{trop}}{c}$$

where $\left\|\bar{D}_i\right\|$ is the true range between user and satellite, and the delay caused by the ionosphere has a negative effect because the physics of the ionosphere advances the carrier. Substituting this equation back into the equation for the phase difference between the received signal phase and the phase generated within the receiver (Equation 4.12), we have

$$\Delta\Phi = -f(\delta_{uc} - d_{sc}) - \frac{f}{c}\left(\left\|\bar{R}_i - \bar{R}_u\right\| + \delta_{ei} - c\delta_{ion} + c\delta_{trop}\right) \qquad (4.13)$$

The term, $\Delta\Phi$, represents total phase. The receiver can only measure the fraction of phase difference. However, the difference in phase between the incoming and generated signals could also have many integer number of cycles as well as the fraction. These integer number of cycles are not directly measured by the GPS receiver and therefore are lost. We can write the phase difference as a sum of three parts:

$$\Delta\Phi = \Phi_{frac} + Int(t_0, t_r) + N(t_0) \tag{4.14}$$

where $N(t_0)$ is the initial integer difference in whole cycles between the received signal phase and the receiver-generated signal phase. With perfect user and satellite clocks, $N(t_0)$ would be the integer number of wavelengths between user and satellite. The term $Int(t_0, t_r)$ represents the number of whole cycles that have passed between the start of navigation and the current time t_r. Carrier-phase receivers are able to keep track of the number of cycles since the initial time.

Since we can physically measure Φ_{frac} and we can keep track of $Int(t_0, t_r)$, then we can consider $N(t_0)$ as the ambiguity term. Let us define the "measured" phase as the sum of the two non-ambiguity terms:

$$\Phi_{meas} = \Phi_{frac} + Int(t_0, t_r) \tag{4.15}$$

Then for total phase we have

$$\Delta\Phi = \Phi_{meas} + N(t_0) \tag{4.16}$$

Substituting this into the carrier-phase equation gives

$$\Phi_{meas} = -f(\delta_{uc} - d_{sc}) - \frac{f}{c}\left(\left\|\bar{R}_i - \bar{R}_u\right\| + \delta_{ei} - c\delta_{ion} + c\delta_{trop}\right) - N \tag{4.17}$$

where I dropped the t_0 off the ambiguity term because, in the advent of loss of lock, the new initial time for the ambiguity term would be the time of loss of lock

4.3 Doppler

The GPS receiver generates a carrier signal internally and beats this signal against the incoming received signal. This beating phenomenon is a signal multiplication resulting in a signal frequency equal to the difference between the incoming and generated frequencies. Range rate is the true time derivative of measured phase or, simply, the change of phase between two closely-spaced times divided by that infinitesimal interval:

$$\Delta t = time\,1 - time\,2$$

$$Range\ Rate \equiv \frac{d\Phi}{dt} = \lim(\Delta t \Rightarrow 0)\left\{\frac{-\lambda\Phi_{meas}(time\ 1) - \lambda\Phi_{meas}(time\ 2)}{\Delta t}\right\} \tag{4.18}$$

28

Substituting in our carrier-phase measurement equation from the previous section gives

$$\frac{d\Phi}{dt} = \lim(\Delta t \Rightarrow 0)\left\{\frac{c(\Delta\delta_{uc} - \Delta d_{sc}) + (\Delta\|\bar{R}_i - \bar{R}_u\| + \Delta\delta_{ei} - c\Delta\delta_{ion} + c\Delta\delta_{trop})}{\Delta t}\right\} \qquad (4.20)$$

where

$$\Delta variable \equiv varible(time\,1) - variable(time\,2)$$

Notice that the ambiguity term disappears since it is common to both times, and the other errors can be significantly reduced if the times are close enough together. Taking Equation 4.20 to the limit (One more time--You know I've always been a dreamer), we have the ideal expression for range rate:

$$\frac{d\Phi}{dt} = c\frac{d\delta_{uc}}{dt} - c\frac{dd_{sc}}{dt} + \frac{d\|\bar{R}_i - \bar{R}_u\|}{dt} + \frac{d\delta_{ei}}{dt} - c\frac{d\delta_{ion}}{dt} + c\frac{d\delta_{trop}}{dt} \qquad (4.21)$$

The user velocity is buried in the third term on the right-hand side of Equation 4.21. To relate this measurement to user velocity, you would have to extract the satellite's contribution to the line-of-sight velocity. Satellite position and velocity can be calculated from ephemeris data found in the data message (the data message is modulated onto the carrier).

Example 4.3

Suppose we live in a one-dimensional world with one satellite and one user receiver placed as follows:

The satellite is moving to the right at 100 meters per second. GPS doppler count at the receiver measures positive 50 cycles over a time interval of .01 seconds and the carrier wavelength is .2 meters. Also, assume the speed of light is 3e8 meters per second, and the following hold true:

$$time_1 - time_2 = .01 \qquad \delta_{uc}(time_1) = \delta_{uc}(time_2) \qquad \delta_{sc}(time_1) = \delta_{sc}(time_2)$$

$$\delta_{ei}(time_1) = \delta_{ei}(time_2) \qquad \delta_{ion} = \delta_{trop} = 0$$

Find the velocity (speed and direction) of the user.

Solution

Equation 4.21 yields

$$50 \; cycles = c(0-0) + \Delta \| R_i - R_u \| + 0 - 0 + 0$$

$$50 = \Delta \| R_i - R_u \|$$

$$50 = \Delta R_i - \Delta R_u$$

$$50 = (100)(.01)/(0.2) - \Delta R_u$$

$$\Delta R_u = 5 - 50 = -45 \; cycles = -9 \; meters$$

$$\frac{dR_u}{dt} = -9/.01 = -900 \; meters/\sec$$

and the user is moving to the left because of the negative user velocity. For this example we dropped the vector notation since the problem has only one dimension. This made going from Step 2 to 3 very simple. For higher dimensions it will not be so simple.

♠

For relatively benign trajectories such as straight and level flight, we could make the time interval large and we would most likely get an accurate measure of LOS velocity. Large time intervals tend to smooth out the noise. On the other hand, during high dynamic maneuvers such as a 6-gee turn or a 1 second roll (common with F-16s), the time interval must be small enough to catch changes in the LOS velocities, but short time intervals cause convergence on noise spikes. You can see there is a trade-off between the size of Δt and measurement accuracy during high dynamics.

4.4 Pseudorange delta-range:

As I told you before, this is not the doppler measurement. Rather, some users calculate this as a substitute for doppler. Whereas the doppler is the change in carrier-phase between two closely-spaced times, the pseudorange delta-range is the change in pseudoranges between two closely-spaced times divided by the time interval:

$$\frac{\Delta PR}{\Delta t} = \frac{PR_i (time \; 1) - PR_i (time \; 2)}{\Delta t} \tag{4.22}$$

Thus, our pseudorange delta-range equation is

$$\frac{\Delta PR}{\Delta t} = \frac{c(\Delta \delta_{uc} - \Delta d_{sc}) + \left(\Delta \| \vec{R}_i - \vec{R}_u \| + \Delta \delta_{ei} + \Delta \delta_{ion} + \Delta \delta_{trop} \right)}{\Delta t} \tag{4.23}$$
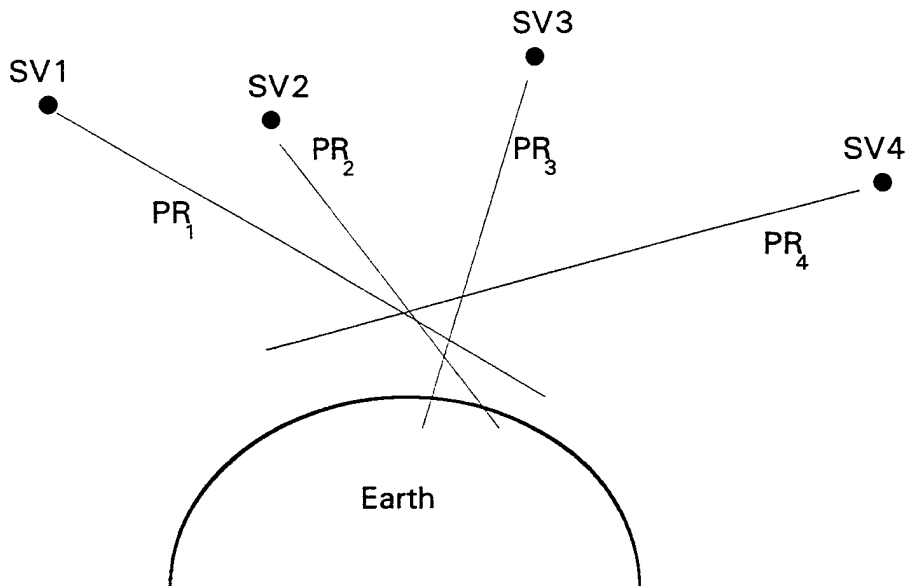
5. *Positioning Using Pseudoranges*

Before we get into the nuts and bolts of the position calculations, I want to point out there are many different types of positioning. For example. some types are :

1. Static positioning
2. Relative positioning
3. Kinematic positioning

One may want to estimate the position of one object relative to the position of another object. An example of relative positioning is the calculation of an approach aircraft's position relative to an airport runway. On the other hand one may want to determine the position of a stationary point with respect to the earth such as a power pole. Such positioning is called static positioning. Also, one may want to maintain track of an object's position as it moves with respect to the earth. This type of positioning is called kinematic positioning. We will concentrate on static positioning in this short course.

5.1 Simple pseudorange equation

Because of the unknown user clock bias in GPS receivers, we'll need at least 4 pseudorange measurements with reasonable geometry. (I'll explain "reasonable geometry" later.) In the following picture I show a 3D problem using 4 GPS satellites. Notice that the position of the user is somewhere "close" to the earth and one could find this position by intersecting spheres with radii equal to the measured ranges plus user clock bias.

So, here are some tidbits of information on GPS that summarize what you should know so far:

1. In each position estimate we have at least 4 unknowns (3 position unknowns and one user clock error). Thus, we need at least 4 equations to solve for position.

2. The user clock error δ_{uc} is the largest error so we include it as an unknown (other errors either ignored, directly calculated, or estimated by some means)

3. A side benefit of GPS is that I can determine how far off my watch is from GPS time.

4. If we ignore all errors except the user clock bias, the pseudorange equation from the previous section becomes

$$PR_i = \sqrt{(x_u - x_i)^2 + (y_u - y_i)^2 + (z_u - z_i)^2} + b \tag{5.1}$$

where I let $b \equiv bias \equiv c\delta_{uc}$ and b has units of distance.

The final navigation parameters must be coordinatized in some frame, like Earth-Centered-Inertial (ECI), Earth-Centered-Earth-Fixed (ECEF), North-East-Down (NED), local navigation system (latitude,longitude,altitude), to name a few. I choose to use the ECEF frame because it is easy to understand, and the other frames can be easily calculated from the ECEF frame by using coordinate transformations such as Direction Cosine Matrices (DCMs).

5.2 Solving for position using pseudoranges

Let's say we have 4 satellites that our receiver is tracking. (Note, it could be more satellites, resulting in an over-determined situation. Later we'll deal with the over-determined situation.) Thus, from above, the 4 measurement equations are

$$PR_1 = \sqrt{(x_u - x_1)^2 + (y_u - y_1)^2 + (z_u - z_1)^2} + b + \textit{other errors}$$
$$PR_2 = \sqrt{(x_u - x_2)^2 + (y_u - y_2)^2 + (z_u - z_2)^2} + b + \textit{other errors}$$
$$PR_3 = \sqrt{(x_u - x_3)^2 + (y_u - y_3)^2 + (z_u - z_3)^2} + b + \textit{other errors} \tag{5.2}$$
$$PR_4 = \sqrt{(x_u - x_4)^2 + (y_u - y_4)^2 + (z_u - z_4)^2} + b + \textit{other errors}$$

Ignoring those other errors gives us 4 unknowns and 4 equations. So, you can solve this directly ....right? Well, you can, but it is very difficult. Notice these equations are nonlinear. It gets even muckier when you have 5 or more measurements if you attempt to solve them directly. Fortunately, the sun is shining and there is an alternative:

Linearize!

32

Yes, you can linearize these nonlinear equations about some nominal point. You'll see that the equations become much more tractable via linearization. You'll also see that the linearization gives us a tool to study geometry effects.

Claim: We usually have a good (albeit approximate) idea where we are on this earth (except for AFIT and TPS students during exam week and AFIT instructors during thesis quarters). For example, I may be lost in the woods in the Jefferson National Forest of Virginia, but I know I'm between 30 and 40 degrees latitude North and 75 and 85 degrees longitude West. My initial guess could be that I'm approximately at 35 N latitude and 80 W longitude. I'm going to call this initial guess the initial "nominal point".

Let the nominal point be given by

$$x_n, y_n, z_n, b_n$$

The difference between the true position/time parameters and the nominal parameters is

$$\Delta x, \Delta y, \Delta z, \Delta b$$

and

$$x = x_n + \Delta x$$
$$y = y_n + \Delta y$$
$$z = z_n + \Delta z$$
$$b = b_n + \Delta b$$

where the true parameters are

$$x, y, z, b$$

Notice, in addition to nominal position, we also have a nominal user clock bias, b_n. Initially, the nominal user clock bias is most likely zero.

Assume we have ECEF coordinates for the satellite (you can calculate ECEF coordinates for the satellites from the ephemeris data). Let's define the nominal pseudorange to satellite i as

$$PR_{ni} = \sqrt{(x_n - x_i)^2 + (y_n - y_i)^2 + (z_n - z_i)^2} + b_n \qquad (5.3)$$

Linearizing Equation 5.2 about the nominals using a first-order Taylor series approximation gives

$$\Delta PR_i \approx \frac{(x_n - x_i)\Delta x}{PR_{ni} - b_n} + \frac{(y_n - y_i)\Delta y}{PR_{ni} - b_n} + \frac{(z_n - z_i)\,\Delta z}{PR_{ni} - b_n} + \Delta b \qquad (5.4)$$

where

$$\Delta PR_i \approx PR_i - PR_{ni} \qquad (5.5)$$

33

Note that ΔPR_1 is a "differenced" measurement.

A look at Equation 5.4 reveals that we have linear equations in the unknowns $\Delta x, \Delta y, \Delta z, \Delta b$. All other quantities we either know or can compute.

Example 5.1

Linearize the following 2D pseudorange equations about nominals x_n, y_n, b_n:

$$PR_1 = \sqrt{(x_u - x_1)^2 + (y_u - y_1)^2} + b$$
$$PR_2 = \sqrt{(x_u - x_2)^2 + (y_u - y_2)^2} + b$$
$$PR_3 = \sqrt{(x_u - x_3)^2 + (y_u - y_3)^2} + b$$

Solution

Define

$$\vec{f}(\vec{v}) = \begin{pmatrix} \sqrt{(x_u - x_1)^2 + (y_u - y_1)^2} + b \\ \sqrt{(x_u - x_2)^2 + (y_u - y_2)^2} + b \\ \sqrt{(x_u - x_3)^2 + (y_u - y_3)^2} + b \end{pmatrix} \equiv \begin{pmatrix} f_1(\vec{v}) \\ f_2(\vec{v}) \\ f_3(\vec{v}) \end{pmatrix}$$

where

$$\vec{v} \equiv \begin{pmatrix} x_u \\ y_u \\ b \end{pmatrix} \qquad \vec{v}_n \equiv \begin{pmatrix} x_n \\ y_n \\ b_n \end{pmatrix}$$

A Taylor series representation gives

$$\vec{f}(\vec{v}) = \vec{f}(\vec{v}_n) + \left[\frac{\partial \vec{f}(\vec{v})}{\partial \vec{v}} \right]_{\vec{v}_n} \delta \vec{v}_n + \textit{higher order terms}$$

Let's pick on the first equation. Ignoring the higher order terms we have

$$f_1(\bar{v}) \approx f_1(\bar{v}_n) + \left(\frac{\partial\left(\sqrt{(x_u - x_1)^2 + (y_u - y_1)^2} + b\right)}{\partial x_u} \right)_{\bar{v}_n} \Delta x$$

$$+ \left(\frac{\partial\left(\sqrt{(x_u - x_1)^2 + (y_u - y_1)^2} + b\right)}{\partial y_u} \right)_{\bar{v}_n} \Delta y$$

$$+ \left(\frac{\partial\left(\sqrt{(x_u - x_1)^2 + (y_u - y_1)^2} + b\right)}{\partial z_u} \right)_{\bar{v}_n} \Delta z$$

$$+ \left(\frac{\partial\left(\sqrt{(x_u - x_1)^2 + (y_u - y_1)^2} + b\right)}{\partial b} \right)_{\bar{v}_n} \Delta b$$

Taking these partials gives

$$\Delta PR_1 = \frac{x_n - x_1}{PR_{ni} - b_n} \Delta x + \frac{y_n - y_1}{PR_{ni} - b_n} \Delta y + \Delta b$$

where

$$\Delta PR_1 \equiv f_1(\bar{v}) - f_1(\bar{v}_n)$$

After first-order analyses on the other equations we have

$$\Delta PR_{2,3} = \frac{x_n - x_{2,3}}{PR_{ni} - b_n} \Delta x + \frac{y_n - y_{2,3}}{PR_{ni} - b_n} \Delta y + \Delta b$$

For neatness, I like to write it all up into a matrix formulation. Assume that we have 4 satellites and write:

$$\begin{bmatrix} \Delta PR_1 \\ \Delta PR_2 \\ \Delta PR_3 \\ \Delta PR_4 \end{bmatrix} = \begin{bmatrix} \alpha_{11} & \alpha_{12} & \alpha_{13} & 1 \\ \alpha_{21} & \alpha_{22} & \alpha_{23} & 1 \\ \alpha_{31} & \alpha_{32} & \alpha_{33} & 1 \\ \alpha_{41} & \alpha_{42} & \alpha_{43} & 1 \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta y \\ \Delta z \\ \Delta b \end{bmatrix}$$ (5.6)

where

$$\alpha_{i1} = \frac{(x_n - x_i)}{PR_{ni} - b_n} \qquad \alpha_{i2} = \frac{(y_n - y_i)}{PR_{ni} - b_n} \qquad \alpha_{i3} = \frac{(z_n - z_i)}{PR_{ni} - b_n}$$

In compact notation, write this as

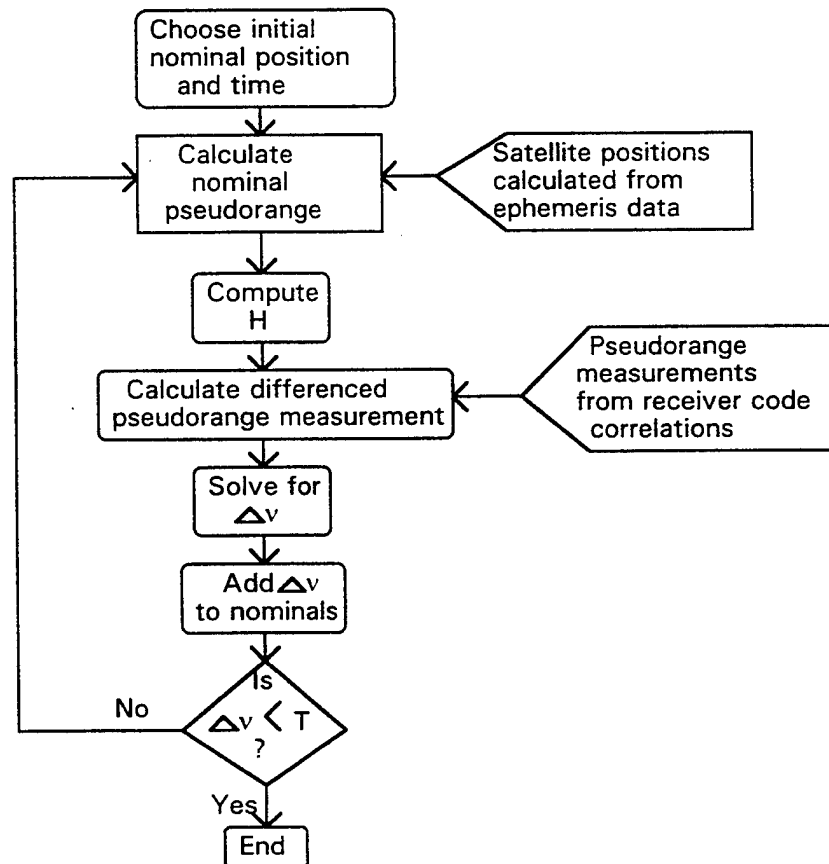$$\Delta P\bar{R} = H\Delta\bar{v} \qquad (5.7)$$

How do you use this equation to solve for the unknowns? Assume for now that the matrix is invertible (more on that one when we talk about geometry). Solving for the unknowns gives us

$$\Delta\bar{v} = H^{-1}\Delta P\bar{R} \qquad (5.8)$$

We can take the result of Equation 5.8 and add it to our nominals to get better nominals:

$$\Delta\bar{v} + \bar{v}_n \Rightarrow \bar{v}_n$$

Recall we started out with nonlinear equations and we linearized them assuming the nominals were reasonably close to the true parameters. If our algorithm is converging properly, these new nominals should be even closer and consequently our assumption even better. This motivates us to try the algorithm again with the new nominals. After a few iterations, we should get to a point where doing more iterations would not make the nominals appreciably better. Notice as the nominals get closer to the real position and time values, the differences, $\Delta x, \Delta y, \Delta z, \Delta b$, get closer to zero. So as a check, the algorithm can monitor the value of the differences and, when a threshold is reached, stop. Next I'll show you a flowchart I use to compute position and time from pseudoranges. Let T be the threshold vector.

```
┌─────────────────┐
│ Choose initial  │
│ nominal position│
│ and time        │
└────────┬────────┘
         ↓
┌─────────────────┐      ┌──────────────────┐
│ Calculate       │←─────│ Satellite positions│
│ nominal         │      │ calculated from  │
│ pseudorange     │      │ ephemeris data   │
└────────┬────────┘      └──────────────────┘
         ↓
   ┌──────────┐
   │ Compute  │
   │   H      │
   └────┬─────┘
        ↓
┌──────────────────────┐   ┌──────────────────┐
│ Calculate differenced│←──│ Pseudorange      │
│ pseudorange measurement│ │ measurements     │
└──────────┬───────────┘   │ from receiver code│
           ↓               │ correlations     │
      ┌─────────┐          └──────────────────┘
      │ Solve for│
      │   Δv    │
      └────┬────┘
           ↓
      ┌─────────┐
      │ Add Δv  │
      │to nominals│
      └────┬────┘
           ↓
      Is Δv < T ?  ──No──┐
           │ Yes
           ↓
        ┌─────┐
        │ End │
        └─────┘
```

Next, I provide a MATLAB algorithm that performs the flowchart above for two dimensions.

```
%LS2d  (Least Squares 2D) Iterates given pseudoranges
   %and satellite positions to give user position solution
%Input is satdat matrix consisting of satellite positions and ranges
%Batch Least squares for 2d problems
ns=size(satdat,2);%ns is the number of satellites
xquess=0;yquess=0;tc=0;%This is our best initial guess
   %of user position and clock bias
iter=5;%number of iterations...could use stopping criterion
sp=satdat([1:2],:);pr=satdat(3,:);
   %Separate satellite position and pseudoranges
qu(1)=xquess;qu(2)=yquess;%Initial guess
%Next, compute nominal ranges from initial guess
   for j=1:ns;
      rn(j)=((qu(1)-sp(1,j))^2+(qu(2)-sp(2,j))^2)^.5;end;
      rn0=rn;
%Next, compute matrix of direction cosines third column
   h(:,3)=ones(ns,1);%It is always ones!
%
%************Start iterations***************
%
for i=1:iter;
   %Compute rest of matrix m for each pass
      for j=1:ns;for k=1:2;
         h(j,k)=(qu(k)-sp(k,j))/(rn(j));end;end;
   dr=pr-(rn+ones(1,ns)*tc);%Compute differenced
      %pseudorange measurements
   dl=inv(h)*dr';%Calculate error states
   tc=tc+dl(3);%Correct old clock bias nominal and
      %get a new nominal
   for k=1:2;qu(k)=qu(k)+dl(k);end;%Compute new nominals
   for j=1:ns;
      rn(j)=((qu(1)-sp(1,j))^2+(qu(2)-sp(2,j))^2)^.5;end;end;
%
%*************End of iterations*************
%
xuser=qu(1);yuser=qu(2);bias=tc;
upos=[xuser yuser]';
```

I leave it up to you to design it for 3 dimensions.

You may wonder if this really saves work as compared to the direct method of solving the non-linear equations. And the answer is yes--if you did your homework, you'd agree! This approach becomes even more attractive as we increase our number of measurements.

For the previous exercise and the development that led up to it, we assumed that we have 4 satellites. It is perfectly reasonable that we would have at least 4 satellites available to us at any given time and place. In fact, with the 24 satellite constellation that we now have, we should theoretically have between 6 and 11 satellites with elevations of 5 degrees or more available to us at any time and at any place on the globe.

Now, suppose we have more than 4 satellites giving valid pseudorange measurements. Do we ignore the extra measurements? No, we shouldn't. Each measurement contains some information about our position. Perhaps I should mention here that some receivers do ignore additional measurements by picking four satellites that give the best geometry. However, adding more measurements can only help the accuracy of our solution. Of course, more than 4 measurements will make the H matrix non-square, and therefore we cannot simply invert the H matrix to find our solution.

What we're going to do next is to deal with having more than 4 measurements of pseudorange. Just for the heck of it, let's say we have 6 measurements:

$$
\begin{bmatrix} \Delta PR_1 \\ \Delta PR_2 \\ \Delta PR_3 \\ \Delta PR_4 \\ \Delta PR_5 \\ \Delta PR_6 \end{bmatrix} = \begin{bmatrix} \alpha_{11} & \alpha_{12} & \alpha_{13} & 1 \\ \alpha_{21} & \alpha_{22} & \alpha_{23} & 1 \\ \alpha_{31} & \alpha_{32} & \alpha_{33} & 1 \\ \alpha_{41} & \alpha_{42} & \alpha_{43} & 1 \\ \alpha_{51} & \alpha_{52} & \alpha_{53} & 1 \\ \alpha_{61} & \alpha_{62} & \alpha_{63} & 1 \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta y \\ \Delta z \\ \Delta b \end{bmatrix} \tag{5.9}
$$

and we want $\Delta x, \Delta y, \Delta z, \Delta b$ as usual. In this case, we have 4 unknowns and 6 equations, clearly an over-determined problem. We can write Equation 5.9 in the form

$$
\Delta P\vec{R} = H\Delta\vec{v}
$$

where H is not square anymore. No $\Delta\vec{v}$ will exactly solve all 6 equations. To solve for the "best" $\Delta\vec{v}$ based on some criterion, we must use a "pseudo-inverse" of H:

$$
\Delta\vec{v} = [H]^{-1} \Delta P\vec{R} \tag{5.10}
$$

Clearly, this matrix equation above represents an over determined situation: we have 4 unknowns and 6 equations. The criterion we will use to find $[H]^{-1}$ will be "least square error". The least square error method will determine a $\Delta\vec{v}_{ls}$ that minimizes:

$$\left(\Delta PR_1 - \Delta P\hat{R}_1\right)^2 + \left(\Delta PR_2 - \Delta P\hat{R}_2\right)^2 + ... + \left(\Delta PR_6 - \Delta P\hat{R}_6\right)^2$$

where $\Delta P\hat{R}_i$ is the value of each row equation using the least square estimates:

$$\Delta P\hat{R}_i = \alpha_{i1}\Delta\hat{x}_{ls} + \alpha_{i2}\Delta\hat{y}_{ls} + \alpha_{i3}\Delta\hat{z}_{ls} + \Delta b_{ls}$$

The least squares equation is as follows:

$$\Delta\hat{v} = \left[H\right]^{-1}\Delta P\bar{R}$$

$$\Delta\hat{v} = \left(H^T H\right)^{-1} H^T \Delta P\bar{R}$$

(5.11)

Note that there are 6 measurements in the illustration above, so the pseudo-inverse of H is $\left[H\right]^{-1}$ and is a 4×6 matrix. You can now add least squares capability to the flow chart given a few pages back. Simply replace Equation 5.8 with Equation 5.11!

Next, we'll look at GPS errors and examine the important role they play.

6. *Measurement Errors*

Two basic types of errors affect the accuracy of the position and velocity solutions offered by GPS: measurement errors and geometrical errors. Measurement errors include errors in each measurement (pseudorange, phase-range, or doppler). These measurement errors come in the form of biases and zero-mean noise. Geometrical errors originate from the geometry of the satellites with respect to each other and to the user. In order to calculate the overall position and velocity solution error, we will find out in the next section that we must multiply the average measurement error by a scalar representation of the geometrical error. This section describes measurement errors while the next section covers geometrical errors.

Recall that we can write the pseudorange measurement from the ith satellite as

$$PR_i = h_i(x_u, y_u, z_u, \delta_{uc}) + \upsilon$$

where υ contains all the error terms except for user clock error (this equation assumes user clock error, δ_{uc}, to be one of the four unknowns). We could just consider υ to be white noise and ignore the fact that many errors are biases. In fact, that's just what we did in Section 5. However, ignoring all errors will put our solution errors on the order of 100 meters even with good geometry. Therefore, in order to improve overall accuarcy we should examine this error term. Some of the most important errors are (note, my list may not be complete!)

$$\upsilon = c\delta_{ion} + c\delta_{trop} + \delta_{ei} - c\delta_{sc} + mp_i + sa_i + rel_i + im_i + cl_i + other\ errors$$

where

$c\delta_{ion} = ionospheric\ error\ (meters)$

$c\delta_{trop} = tropospheric\ error\ (meters)$

$\delta_{ei} = ephemeris\ errors\ (meters)$

$\delta_{sc} = satellite\ clock\ errors\ (meters)$

$mp_i = multipath\ errors\ (meters)$

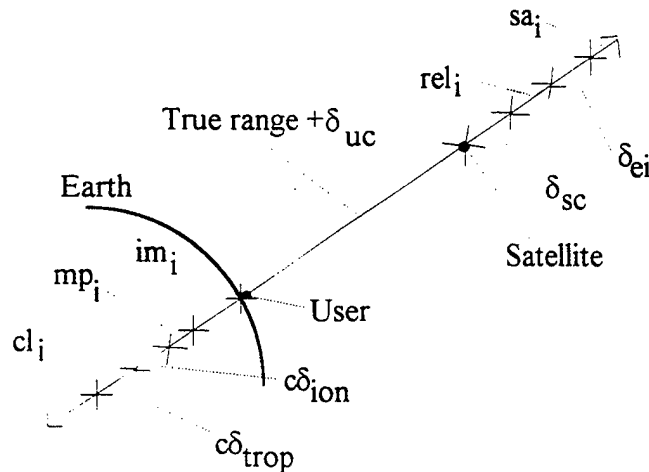$sa_i = selective\ availability\ (meters)$

$rel_i = error\ due\ to\ relativity\ (meters)$

$im_i = error\ due\ to\ antenna\ image\ (meters)$

$cl_i = code\ loop\ errors\ (meters)$

The phase-range measurement in Section 4 contains these same error terms above plus an ambiguity term, and the doppler measurement contains rates of these error terms.

All these errors add to or detract from the pseudorange measurement. The following picture illustrates this:

True range $+\delta_{uc}$

sa$_i$

rel$_i$

δ_{ei}

δ_{sc}

Satellite

Earth

im$_i$

mp$_i$

cl$_i$

User

$c\delta_{ion}$

$c\delta_{trop}$

The pseudorange consists of the sum of all these errors. Note, despite my figure, some errors will subtract and some will add from the true range.

How bad are these errors and what can we do about them? The following table offers a initial attempt to answer this question. In the table, I've divided the errors into three major categories: control/space, user, and propagation link. The third column in the table represents typical magnitudes of errors directly added to the pseudorange measurement.

Category	Error	Typical magnitudes	What can we do?
Control and space segments	1. Ephemeris 2. Satellite clock 3. Selective Availability (SA)	1. 10 meters maximum w/o SA 2. 3e5 meters uncorrected and 1 m corrected w/o SA 3. 100 meters	1. Control Corrects daily, usually ignored 2. Satellite clocks must be corrected, could filter residual 3. Get it turned off!
User segment	1. User clock errors (treated as unknown) 2. Code and carrier loop errors 3. Reciever noise (treated as part of v)	1. 1-100 meters, depends on oscillator 2. 1 to 10 % of code or carrier cycle 3. Depends on receiver	1. Just use a good clock! 2. Usually tolerated, filtering may help 3. Filtering ("other errors" in equation)

| Propagation link | 1. Ionospheric refraction
2. Tropospheric refraction
3. Multipath
4. Imaging
5. Relativity effects | 1. 50 to 150 meters
2. 2 to 20 meters
3. Depends on antenna location
4. Also depends on antenna location
5. A few centimeters at maximum | 1. Model, ignore, or use dual frequency
2. Model, or ignore
3. Beam shape, or just place your antenna better!
4. Better antenna material, beam shape
5. Usually ignored |

Many of these errors will be cut out completely or almost completely by using differencing techniques. Nonetheless, differencing techniques may be impossible in many applications. Next, we examine each of these errors terms one by one, starting with the most damaging and proceeding to the least.

6.1 Selective Availability (SA)

Yes, believe it or not, selective availability is the worst, and to think that it is the only "error" intentionally set! The goal of SA is to deny full GPS accuracy to unauthorized users (the U.S. Department of Defense determines authorized users). SA degrades horizontal position accuracy to 100 meters and vertical position accuracy to 156 meters (95% probability level). It also degrades velocity to 0.3 meter per second and time to 340 nanoseconds. SA does not degrade authorized users. One day soon I believe SA will be turned off; therefore, this section will be brief.

SA spoils GPS accuracy in two ways. We call the first method the δ-process and it consists of dithering the satellite clock frequency. The second method is the ε-process in which the Keplarian parameter data words describing satellite orbits are truncated. These two methods have different characteristics and the DOD can turn one or both of them on or off at will.

Having an accurate satellite clock is vital to obtaining accurate pseudoranges. The δ-process dithers the fundamental clock frequency with an amplitude of about plus and minus 2 hertz and a period of 5 to 10 minutes. A satellite frequency dither directly affects measurements by increasing measurement error. For instance, the δ-process causes 10 to 50 meters of pseudorange error, ultimately causing a similar error in user position estimation. Without SA, satellite clock error produces only about 1 meter error in pseudorange.

The ε-process truncates the ephemeris data so users cannot calculate precisely the position of the satellite. Recall from Chapter 4 that we must know the satellite location in order to compute user position. Not knowing precisely the locations of the satellites will negatively impact user position estimation accuracy. These truncations cause satellite position errors with periods of 2 to 8 hours and magnitudes of 50 to 100 meters. Note, satellite

position errors translate almost directly to user position estimation errors. Over the span of 2 to 8 hours, the ε-process will tend to circle the true position.

How do we circumvent the SA problem? One method includes the use of differencing techniques. For example, using a common satellite and two different receivers and subtracting the measurements will cancel SA effects as long we synchronize the measurements. However, we must use a reference station which may not be available, and data rates between the reference and user must be high enough to overcome changes in SA. Another method includes an attempt to model SA. For example, we may describe SA as an nth-order markov process with certain parameters. Of course the best way to circumvent SA would be to become an authorized user. Authorized users can decrypt correction data available in the ephemeris and add that data to the part of the message available to everyone. In this way, SA effects are cancelled.

6.2 Ionospheric Errors

Generally, we divide the atmosphere into 3 major layers: the troposphere, the stratosphere, and the ionosphere. The ionosphere spans from about 50 to 100 kilometers in altitude (this range depends on the scientist as well as real physical processes). The ionosphere differs from the other two layers in many ways, sometimes bizaar ways. The biggest difference comes from the fact that the ionosphere disperses electromagnetic signals in the L-band region of the spectrum (this region is home to GPS signals).

Dispersion of electromagnetic radiation (such as GPS signals in the ionosphere) stem from diffraction and refraction of the signals. The ionosphere is a dispersive median. Ultraviolate radiation from the sun creates ions by stripping the upper atmosheric atoms of their electrons. Consequently, there are multitudes of free electrons and ions running around over our heads. We call a gas that contain ions and free electrons a plasma. A plasma has many interesting properties, some of them down-right weird. One property is signal dispersion. In other words, the velocity (and thus the index of refraction) is a function of frequency. The ionosphere disperses a signal with closely spaced frequencies like GPS.

A single waveform with one frequency travels through the ionosphere with a velocity called phase velocity. A group of waveforms with closely spaced frequencies travel through the ionosphere with a velocity called group velocity. In GPS, we must deal with group velocity when using code-phase, and phase velocity when using carrier-phase. Without getting you lost in a bunch of algebra (see my text for details), the biases caused by the ionosphere are

$$\Delta L_p = \left(1 - 1 + \frac{\alpha_1}{f^2}\right)L = \frac{\alpha_1}{f^2}L = \frac{40.3NL}{f^2} \tag{8.1}$$

$$\Delta L_g = \left(1 - 1 - \frac{\alpha_1}{f^2}\right)L = -\frac{\alpha_1}{f^2}L = \frac{-40.3NL}{f^2} \tag{8.2}$$

43

The length of the path through the ionosphere denoted by L has units of meters, and f represents the frequency of the carrier. The quantity NL is the total electron count in a column along the path with a cross sectional area of one meter. Sometimes you'll see TEC (total electron count) for NL.

To find TEC, simply count all the free electrons you see between you and the satellite. What, you can't? Estimating TEC is difficult. Researchers have devised multitudes of ways to measure TEC along a vertical path relative to the user (VTEC for vertical total electron count). Simple trigonometry gives the following approximation to VTEC:

$$VTEC \approx \sin(e)TEC \tag{8.3}$$

where e represents the elevation angle of the satellite with respect to the user. Of course, Equation 8.3 could be a very bad approximation since the path of the signal and this vertical column exist in different areas of the ionosphere. Using Equation 8.3, the path length changes for carrier-phase and code-phase are

$$\Delta L_p = \frac{40.3VTEC}{\sin(e)f^2}$$

$$\Delta L_g = \frac{-40.3VTEC}{\sin(e)f^2} \tag{8.4}$$

where carrier-phase corresponds to phase velocity and code-phase (pseudorange) corresponds to group velocity.

VTEC ranges from 1e16 to 1e18 electrons per meter squared and depends on many factors. These factors include sunspot activity with the 11-year cycle (VTEC is highest at solar maximum), the seasons (VTEC is higher in winter), the time of day (VTEC is highest around 1400 hours and lowest between midnight and 6:00 am), and the location of the user. The worst cases depicted in terms of range errors are as follows

Zenith	50 meters
Horizon	150 meters
Night	20% of the day
Annual	November is 4 times July
Sunspot	Maximum is 4 times minimum

Researcher's work in this area continues to this day, and will probably continue for a long time as long as we understand so little about the ionosphere.

Ionospheric delay computation with dual frequencies

44

The fact that Equation 8.4 depends on frequency gives us a great opportunity to find ΔL without worrying about VTEC. Two carrier frequencies, L1 and L2, exist for GPS. The path length change using carrier L1 differs from the path change using carrier L2. Let's pick on pseudoranges, so we'll use the delay associated with group velocity in Equation 8.4. Pseudorange measurements from L1 and L2 yield the following:

$$PR_{L1} = true\ range - \frac{\alpha_1}{f_{L1}^2} + all\ other\ errors$$

$$PR_{L2} = true\ range - \frac{\alpha_1}{f_{L2}^2} + all\ other\ errors$$

(8.5)

Subtracting the two pseudoranges gives us an expression for calculating α_1 without worrying about TEC:

$$PR_{L1} - PR_{L2} = \tau = \frac{\alpha_1}{f_{L1}^2} - \frac{\alpha_1}{f_{L2}^2}$$

$$\alpha_1 = \frac{\tau}{\dfrac{1}{f_{L1}^2} - \dfrac{1}{f_{L2}^2}}$$

(8.6)

This method of calculating ionospheric error proves extremely effective -- that's one reason P-code capable receivers are so great -- they allow us to measure code-phase on L2 as well as L1.

6.3 Tropospheric Errors

The troposphere and stratosphere extend to 50 and sometimes as high as 80 kilometers high. Scientists call these two layers the "neutral" atmosphere because no dispersion exists as it does in the ionosphere (for frequencies less than about 20 GHz). Thus, in this neutral zone of the atmosphere, group and phase velocities are equal since refraction does not depend on frequency. This makes for good news and bad news. The good news: modelling a non-dispersive media is easier than modelling a dispersive media. The bad news: with no frequency dependent terms, we cannot back out the path length error by using dual frequencies like we did for the ionosphere. From now on, I'll refer to the entire neutral zone as simply the troposphere to be consistent with the literature.

Tropospheric refraction contains two basic components: a dry and a wet component. For vertical signal paths, the dry component causes about 9 times the error contributed by the wet component. However, both components depend heavily on elevation angle -- tropospheric refraction increases as the satellite approaches the horizon. We can easily and accurately calculate the dry component by just measuring the surface pressure and elevation of the satellite. On the other hand, the wet component depends on atmospheric conditions all along the signal path. Atmospheric conditions include water content, temperature, air pressure, altitude of user, and elevation of the satellite.

A myriad of models...

Multitudes of tropospheric models exist, mostly depending on how one measures or calculates the wet component. Next, I'll present a few of these models but realize that these are just a smattering of all the available models you can choose from -- what the heck, you could make up your own model and become famous!

Saastamoinen model -- easy for you to say

I'll do no derivations here, I'll just give you the straight scoop. The Saas (short for Saastamoinen) model is

$$\Delta L = \frac{.002277}{\cos z}\left\{ p + \left(\frac{1255}{T}+.05\right)p_w - \tan^2 z \right\}$$ (8.7)

where ΔL denotes the path length change in meters due to the troposphere. (Note, the troposphere always causes a delay, never an advance, so ΔL positive means delay in this case.) The zenith angle is z (or $90^\circ - e$), p is the atmospheric pressure in millibars, T is the temperature in Kelvin, and p_w is the pressure in water vapor in millibars.

Ashjaee model

I call the following model the Ashjaee model because Ashjaee uses it in one of his papers (my source may be wrong, if it is, my apologies!). The model is as follows:

$$\Delta L = \frac{2.4225}{\sin e + .025}$$ (8.8)

You can see the simplicity of this model, depending only on the elevation angle of the satellite. The delay given by the Ashjaee model agrees closely with the delay given by the Saas model.

Hopfield model

Back in 1969 a guy named Hopfield used real data to model the dry refractivity of the troposphere as a function of altitude. Since those days, researchers have constructed many variations of the Hopfield model, including models of the wet component. Like the Saas model and unlike the Ashjaee model, the basic Hopfield model contains both dry and wet components:
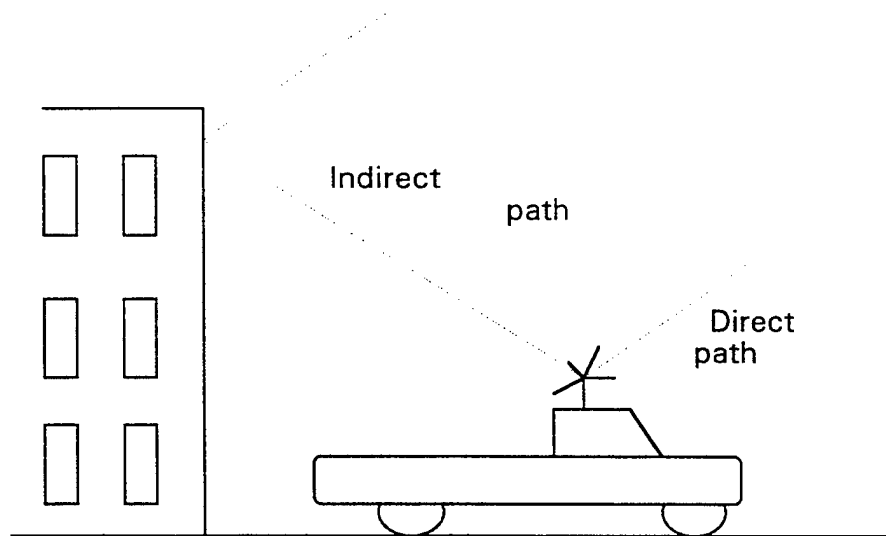
$$\Delta = \Delta_d + \Delta_w$$

where

$$\Delta_d = \frac{.000015528\frac{p}{T}}{\sin\sqrt{e^2 + 6.25}}\left(148.72T - 488.3552\right)$$

$$\Delta_w = \frac{p_w\left(817.96 - .028512T\right)}{T^2 \sin\sqrt{e^2 + 2.25}} \tag{8.9}$$

Again, the wet and dry components agree well with the Saas model and the total delay agrees closely with all 3 models.

6.4 Multipath Error

Signals from GPS satellites can reflect off objects near the receiver and/or satellite. Consequently, these reflected signals follow multiple paths between the satellite and the receiver. Paths other than the line-of-sight direct path will have longer transit times, and therfore longer range measurements. When the receiver picks up these indirect signals, we have multipath error. The following figure illustrates my chevy truck next to a tall building with a GPS antenna on the truck's roof. The signal following the indirect path will indicate a larger pseudorange, an increase in error possibly on the order of tens of meters. I show only one indirect path, but there could be others. For example, another indirect path would include a path from the satellite to the hood of my truck to the antenna.

The situation-dependency of multipath makes it difficult to overcome. However, some improvements and fixes exist. Obviously, the best thing we could do is to locate our antenna away from reflective surfaces. Aside from that, we could use an antenna with higher gains toward the direct signal. Also, our antenna could be such that it rejects left-handed circularly polarized signals. GPS satellites transmit right-handed circularly polarized signals, and so reflected GPS signals are left-handed circularly polarized. Signal processing techniques offer

47

promising methods to overcome multipath, especially in the frequency domain. Finally, in some situations, one may be able to average out multipath effects if those effects change sufficiently.

6.5 Antenna Phase Center Movement

Another source of error related to multipath is the unintentional movement of the phase center of the antenna. Nearby refective surfaces, like that building beside my truck, may have an image of the antenna and move the phase center of the real antenna to somewhere between the antenna and the reflective surface. Consequently, position estimates are with respect to some other point other than the cab of my truck. Also, nearby objects may become the antenna. For example, the skin of my truck may become the antenna. In my case, that wouldn't do much harm, but if I parked near a 40-acre building, who knows where the phase center would be! Sometimes a certain manufacturer's antenna may have a phase center separated from the physical antenna. Also, this phase center may change as the antenna rotates. One common fix for phase center movement other than re-location is beam shaping.

6.6 Ephemeris Errors

As we saw in previous chapters, the positions of the satellites at given times are not precisely known because of perturbations. Master control periodically sends corrections to the satellites, which ultimately come to us in the data message. However, we still have to deal with residual ephemeris errors. Example 8.2 illustrates typical sensitivity of solution error to emphemeris error. In the near future, especially with laser tracking capability of the Block 2a satellites, residual ephemeris errors will be less than one meter.

Three methods exist to extract the position error of a satellite. First, we could model the satellite position and estimate it along with the other unknowns. One way to estimate satellite position is to use other navigation systems and filter the information (for example, use a Kalman filter in an integrated approach). Secondly, we could use differential techniques between multiple receivers and a common satellite. In this way, errors in satellite position cancel out. Lastly, we could ignore ephemeris errors. In most situations satellite position errors are too small to matter.

6.7 Satellite Clock Error

As we discussed in Chapter 6, we remove a large portion of the satellite clock error by calculating the *delt* parameter from data given to us in the data message. However, small residual clock errors still lurk. Example 8.1 illustrates typical sensitivity of solution error to satellite clock error (in the form of pseudorange error). All my comment pertaining to ephemeris errors above also pertain to satellite clock error.

6.8 Code Loop and Carrier Loop Errors

Code loop and carrier loop errors depend on the quality of the receiver. Typically, recievers correlate internally-generated phase with received phase with an accuracy between 1 and 10 percent. Therefore, 300-meter C/A code chips result in code loop errors between 3 and 30 meters. For P-code, 30-meter code chips result in code loop errors between .3 and 3 meters. For carrier-phase, 20-centimeter cycles result in carrier loop errors between 2 millimeters and 2 centimeters. Note that differential GPS elliminates or significantly reduces all errors except code loop and carrier loop errors. That is why DGPS has errors on the order of 10 meters (C/A-code) and 1 meter (P-code), and carrier-phase GPS has errors measuring in centimeters!

6.9   Relativity Effects

Special and general relativity affect GPS in a number of ways. These effects are due mainly to the gravitational field of the earth, the motion of the earth, and the motion of the satellite. The gravitational field causes perturbations in the satellite orbit, curves GPS signals through space-time curvature, and dilates time between satellite and user. Assuming the center of the earth is an inertial frame (not really, but close enough), the stationary user travels with respect to the inertial frame with a speed of about 25000/24 times cosine of latitude miles per hour. Also, a GPS satellite travels approximately 2.5 miles per hour with respect to the inertial frame. The difference in velocities between the user and satellite causes dilation in time, mass, and length.
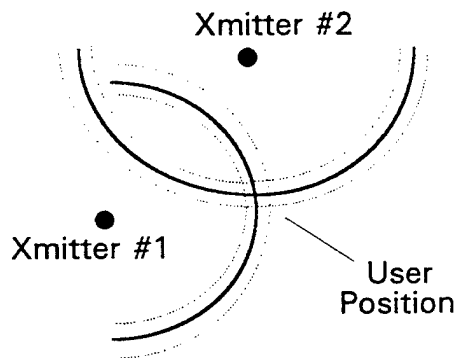
Although relativity affects GPS in many ways, by far the greatest affect is the apparent change in frequency between satellite and user. The frequency shift comes about because of both special and general relativities. Special relativity arises from relative velocities and general relativity arises from accelerations (including gravitation). Recall that time is the inverse of frequency. Both the relative difference in gravitational potential (general relativity) and the relative difference in velocities (special relativity) dilate time, and therefore dilate frequency. Fortunately, the designers of GPS compensated for this frequency difference by offsetting the satellite clock to read 10.22999999543 MHz in order to make the received frequency of the P-code chips to be exactly 10.23 MHz.

7. Geometric Analysis

Each pseudorange measurement has ranging errors associated with it. Clearly, these errors on the pseudorange measurements will translate directly into position estimate errors. However, this was only half of the story. Now you'll hear the rest of the story...."Page two!"
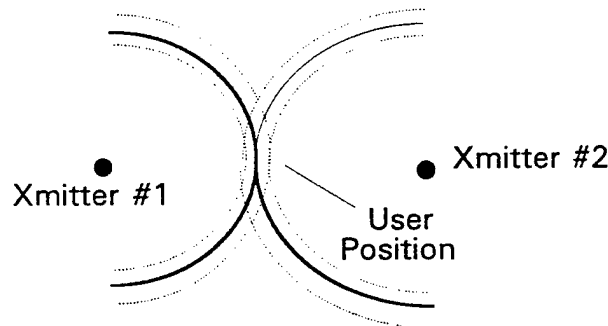
Geometric Errors:

Geometry affects the accuracy of position estimates when ranging techniques are employed (like GPS). Look at the picture below. (Aha, made you look!)

We will assume passive radionavigation with synchronized transmitter and user clocks. The envelops formed by the two dotted line pairs in the picture represent the ranging measurement errors. The situation above obviously has good geometry since the user can nail his position somewhere in the square-like shape arising from the intersection of the two dotted line pairs.

Now consider the situation where the second transmitter is located on the opposite side as the first transmitter:

Again, the measurements have the same errors as before but the position estimate is much worse. In fact the best we can do is to place the user's position within the elongated diamond-like shape described by the intersection of the two pairs of dotted lines. Essentially we have lost one dimension of information on user position by placing the transmitter in such a dumb

spot. We know the left-right position of the user fairly well, but we do not know much about the user position along the line from bottom to top of the paper.

Clearly geometry of the satellites (by geometry of the satellites I mean their relative positions) plays an important role in the accuracy of navigation, just as did the errors on the measurements. This should be clear from the simple 2D examples above.

Next let's find the special scalar number that indicates the quality of satellite geometry. Recall from Equation 5.31 we had

$$\Delta \vec{v} = H^{-1} \Delta P \vec{R} \tag{7.1}$$

Equation 7.1 represents a linearized set of relations. The non linearity of this problem forced us to iterate on the solution beginning with the initial nominal position, and ending the iterations with our position estimate. The matrix H depends on our estimate, therefore H evolves throughout the iterations. In the subsequent text I will assume that we have iterated our algorithm, and so the parameters I use for the geometry analysis come from my final position estimate.

The problem at hand is to quantify the error in the left-hand side of Equation 7.1 (solution error, or position error in this case) by quantifying geometrical and measurement errors in the right-hand side. Before, we found that the measurement error contribution comes from the $\Delta P \vec{R}$ term. Intuition tells us that the geometrical contribution comes from the H term. As we will see, that intuition is correct.

After much algebra, sweat, and tears, we can find that the root mean square (RMS) of the position error is related to the one-sigma of the measurement error (assuming all measurement errors are the same) by

$$\sigma \sqrt{Trace\left[\left(H^T H\right)^{-1}\right]} \tag{7.2}$$

The σ term of Equation 7.2 is the User Equivalent Ranging Error (UERE) contribution to the RMS error while the remaining term is the geometrical contribution to the error. A common name for the geometrical contribution to the error is "Geometrical Dilution of Precision", or GDOP. Using the GDOP notation we have

$$RMS_{\Delta v} = \sigma * GDOP$$
$$GDOP \equiv \sqrt{Trace\left[\left(H^T H\right)^{-1}\right]} \tag{7.3}$$

Another form of Equation 8.3 is

$$GDOP = \sqrt{V_x + V_y + V_z + V_t} \tag{7.4}$$

where V_i for $i = x,y,z,t$ are the diagonal components of the matrix $\left(H^T H\right)^{-1}$.

A few comments are due on Equation 7.3. First, we now have a scalar measure relating geometry of the satellites to the strength of the position/time estimate errors. From Equation 7.3 we see that the higher the GDOP, the worse the solution error. Also, GDOP multiplies the UERE term. Finally, Equation 7.3 gives us a convenient method to calculate GDOP. We already have the H matrix--that's the direction cosine matrix we used in our least squares iterative algorithm (but be careful, here we are using only 4 satellites, not more).

Next, I will list properties of GDOP:

1. GDOP ranges from about 1.5 to huge numbers. A GDOP of about 3 is average. Thus if we have pseudorange errors of 10 meters then position errors will have an RMS value of 30 meters.

2. The best GDOP occurs when the satellites are spread out in all 3 dimensions. For example, one satellite overhead and 3 satellites spread 120 degrees apart close to the horizon is the ideal satellite geometry. Poor GDOP occurs when satellites are bunched together, or all are close to the horizon, or all are in a line. In other words, bad GDOP comes about from the loss of at least one spatial dimension (all satellites in a plane has lost one dimension while all satellites in a line has lost two dimensions).

3. GDOP is independent of the coordinate system. It doesn't matter whether we use the ECEF frame defined in this text or some other earth frame or even an inertial frame. GDOP only depends on the relative position of the satellites. Of course the individual components of the H matrix will change--these are the direction cosines--but the term

$$\sqrt{TR\left\{\left(H^T H\right)^{-1}\right\}}$$

will remain the same.

4. GDOP was used to design the current GPS constellation. Jorgensen's article presents case studies of various GPS constellations but he did not present the 6-plane case which was eventually used for GPS.

5. GDOP is used by some receivers to pick the best four visible satellites. We already know how to calculate GDOP once we iterate and obtain a good position/time estimate. Conceivably, receivers could use this GDOP calculation for each combination of 4 visible satellites to pick which four gives the best geometry. There is a much better way to do this without having to calculate position/time for each satellite foursome (using the volume of a tetrahedron for example).

Other measures of dilution of precision (DOP) exist besides GDOP. GDOP includes both three-dimensional position and time (4 variables), but there may be situations where we

are only concerned about a subset of the four variables. Below I list the 4 most common DOPs besides GDOP.

1. PDOP is the dilution of position only. To calculate PDOP, use Equation 7.4 without the time term:

$$PDOP = \sqrt{V_x + V_y + V_z} \tag{7.5}$$

Average PDOP is about 2.6.

2. HDOP is the dilution of position in the horizontal plane only. To calculate, use Equation 7.4 without the time and z terms:

$$HDOP = \sqrt{V_x + V_y} \tag{7.6}$$

Average HDOP is about 1.45.

3. TDOP is the dilution of time only. To calculate, use only the time component of Equation 7.4:

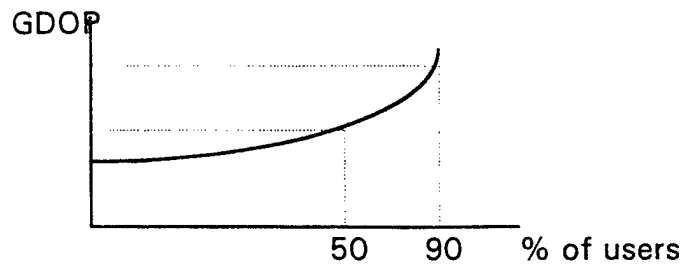$$TDOP = \sqrt{V_t} \tag{7.7}$$

Average TDOP is about 1.2.

4. VDOP is the dilution in vertical position only. To calculate, use only the vertical component of Equation 7.4:

$$VDOP = \sqrt{V_z} \tag{7.8}$$

Average VDOP is about 2.5.

An important observation of Equations 7.5 through 7.8 is that vertical dilution of position is much worse than the others. This happens because all the satellites are generally above the users. Horizontal geometry is good because satellites can be on all sides of the user. Geometry would be improved in the vertical direction if somehow satellites could be placed *below* the user as well. Later we will comment on the use of pseudolites for this very purpose (pseudolites are ground stations that are used like GPS satellites).

GDOP is not distributed as gaussian, so instead of using distributions or density functions we usually represent the statistics of GDOP in percentiles. For example. I may say that 50% of GPS users fall under a GDOP of 2.6. Or 90% of GPS users fall under a GDOP of 3.5 and so on. We could plot curves that look like the following:

GDOP ... 50 90 % of users

We can see from this curve that regardless of what we do, somebody somewhere may have bad GDOP (in other words, those who happen to be on the right-hand side of the plot). Lately, since at least 24 GPS satellites are in orbit, one can theoretically have great GDOP anytime anywhere. In the not so good old days (late 80's and early 90's) we had a much smaller number of orbiting GPS satellites and so GDOP was a major factor. You should also note that modern users can also have bad GDOP if some satellites cannot be picked up by the receiver (from shading and such) or if the receiver picks a bad combination of satellites.

That brings us to that disgusting topic of *outages*. Outages are those situations when we cannot determine a reasonable position/time solution due to the relative positions or lack of visible satellites. There are two causes of outages:
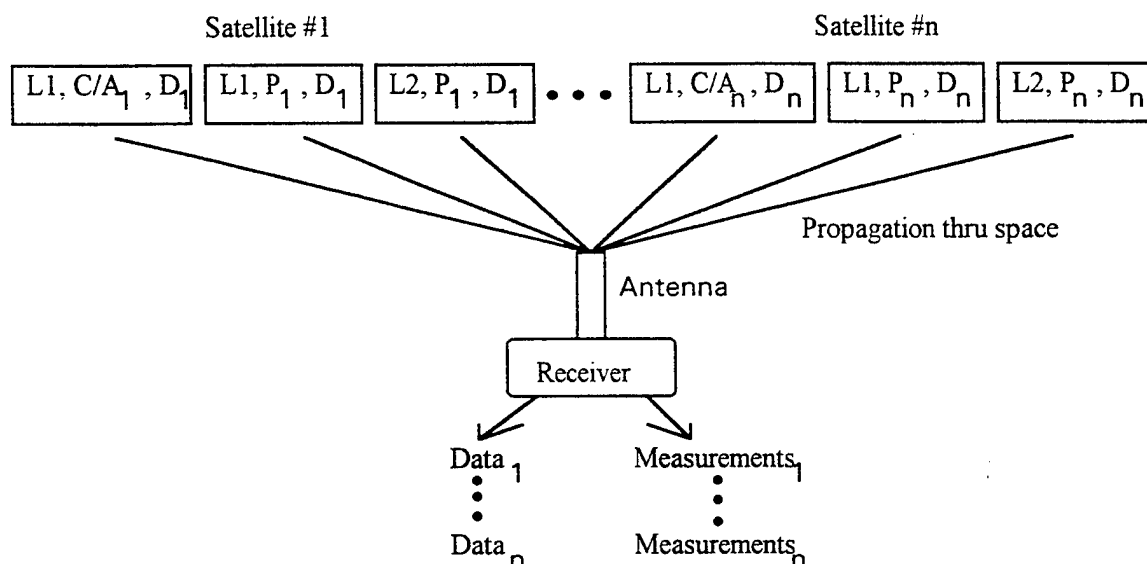
1. Less than 4 satellites in view. Theoretically, there are always 6-11 satellites above a 5 degree mask angle. However, users could shade themselves from satellite reception due to trees, buildings, a wing of the aircraft, tunnel, etc.

2. Poor GDOP. You can receive signals from a zillion satellites, but if they are all bunched together or strewn out in a line, you can forget about a reasonable navigation solution.

Note, for 2-D navigation (latitude and longitude only), outages are less likely since we would only need 3 pseudoranges to solve for 2-D position and time. Also, integration of an inertial navigation system or other navigation systems like LORAN with GPS would help the navigator to get through outages. One could employ a Kalman filter and model motion of the user so that during outage conditions the user can rely more on the propagation equations and less on the update equations. Another solution to the outage problem is to use a precise clock (see the Sturza article for an example). Precise clocks enable the receiver to get a 3-D position fix with only 3 pseudoranges since precise clocks do not drift very much.

8. The Receiver, Signal Structure, and Data Message

So far we have come from the systems point of view in our study of GPS. We've discussed how to compute user position and velocity using the GPS measurements available from the receiver. Given reasonable measurements, we can manipulate the data to be useful to us. Now for this section, let's examine GPS from the signal processing point of view. What do GPS signals look like? How are they generated at the satellites and how are they received by the receivers? What do the guts of the receiver look like? How does the receiver compute pseudorange and carrier-phase?
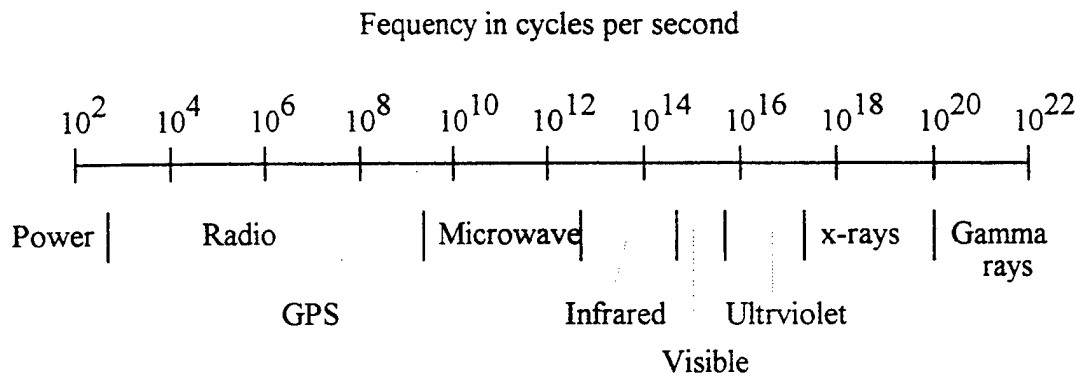
A GPS signal contains carrier, code, and a data message. The satellite modulates the carrier with code and data and transmits the signal towards the earth. The receiver picks up this signal, demodulates it, and extracts the data and measurements. Each satellite imposes two codes, C/A and P, on two frequencies, L1 and L2. The following figure illustrates this process.

Satellite #1   Satellite #n

| $L1, C/A_1, D_1$ | $L1, P_1, D_1$ | $L2, P_1, D_1$ | $\bullet\bullet\bullet$ | $L1, C/A_n, D_n$ | $L1, P_n, D_n$ | $L2, P_n, D_n$ |

Propagation thru space

Antenna

Receiver

Data$_1$ Measurements$_1$
$\bullet$ $\bullet$
$\bullet$ $\bullet$
$\bullet$ $\bullet$
Data$_n$ Measurements$_n$

Note that L2 only carries P-code. For the remainder of this section, we split the topic into two subsections: signal generation at the satellite and signal processing at the receiver.
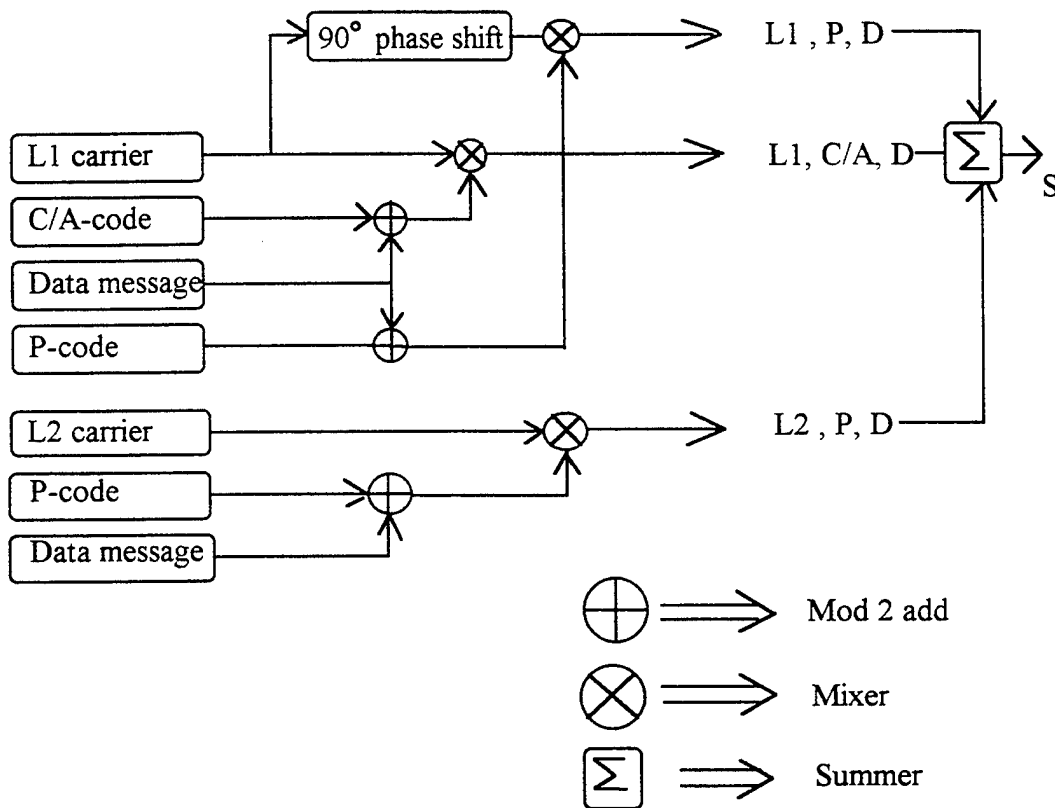
8.1 Signal Generation at the Satellite

First, let's view GPS carrier frequencies within the context of the whole electromagnetic spectrum. L1 carrier propagates at 1575 MHz and L2 at 1227 MHz. It's better to show a picture than for me to babble on about it, so here it is:

Fequency in cycles per second

$10^2 \quad 10^4 \quad 10^6 \quad 10^8 \quad 10^{10} \quad 10^{12} \quad 10^{14} \quad 10^{16} \quad 10^{18} \quad 10^{20} \quad 10^{22}$

Power | Radio | Microwave | | | | x-rays | Gamma rays

GPS   Infrared   Ultrviolet

Visible

GPS signals reside in the lower end of the microwave band. This region of the spectrum limits us to line-of sight in our atmosphere (except when reflecting surfaces cause multipath errors).

At the GPS satellite, the carrier, code, and data are combined in a certain way. Again, the following picture will be worth 953 of my words.

90° phase shift

L1 carrier

C/A-code

Data message

P-code

L1 , P, D

L1, C/A, D

L2 carrier

P-code

Data message

L2 , P, D

Σ → S

⊕ ⟹ Mod 2 add

⊗ ⟹ Mixer

$\boxed{\Sigma}$ ⟹ Summer

Note that the data and codes are modulo 2 added while the data/code combinations are mixed (*i.e.*, modulated) with the carriers. Also the C/A-code is shifted 90 degrees in phase (phase quadrature) with respect to P-code for L1. If the carrier modulated with C/A code is a sine function, then the carrier modulated with P-code is a cosine function. Cosine and sine functions are orthogonal on the interval $0 \le t \le 2n\pi, \quad n = 1,2,3\cdots$:

$$\int_0^{2n\pi} \cos(w_1 t + \phi)\sin(w_1 t + \phi)dt = 0$$

Therefore the C/A-code and P-code portions of the L1 signal are orthogonal. Denoting

$$P_i \equiv P \text{ code for ith satellite}$$
$$G_i \equiv C / A \text{ code}$$
$$D_i \equiv Data \text{ message}$$
$$w_1 \equiv L1 \text{ radian frequency}$$
$$w_2 \equiv L2 \text{ radian frequency}$$
$$A_p, A_c, A_b \Rightarrow scalar \text{ amplitudes}$$
$$\phi_p, \phi_c, \phi_b \Rightarrow signal \text{ phases}$$

we have the following formula for the signal being transmitted from the *i*th satellite:

$$S_i = A_p P_i D_i \cos(w_i t + \phi_p) + A_c G_i D_i \sin(w_i t + \phi_c) + B_p P_i D_i \cos(w_2 t + \phi_b) \qquad (8.1)$$

The scalar amplitudes are related to each other as:

$$A_c = 2A_p = 4B_p$$

The received energy at the receiver is typically

$$C / A \text{ code} \quad on \quad L1 \Rightarrow -160 \quad dbwatts$$
$$P \text{ code} \quad on \quad L1 \Rightarrow -163 \quad dbwatts$$
$$P \text{ code} \quad on \quad L2 \Rightarrow -166 \quad dbwatts$$

The data and codes contain bits and chips with values of plus and minus ones:

The satellite combines the data and codes by modulo two addition:

$$Data \oplus code \Rightarrow 1 \quad \textit{bit value = chip value}$$
$$\Rightarrow -1 \quad \textit{bit value} \neq \textit{chip value}$$

The data/code mixes with the carrier using binary biphase modulations. When the data/code switches from -1 to 1 or from 1 to -1, the phase of the carrier switches 180 degrees.

GPS uses a technique called spread spectrum. Plotting the energy versus frequency of random noise (same as "white" or "totally random noise") results in a flat spectrum. This means that random noise is uncorrelated from one moment to the next. GPS codes are not totally random, but they are close, and so are called "pseudo-random noise" codes. Modulating a carrier with data alone (no codes) causes just a little spreading in frequency. For GPS this spreading due to the data alone would be plus and minus 50 Hz, or a 100 Hz spread. On the other hand, pseudo-random noise codes spread the signal much further. GPS C/A codes spread the signal by plus and minus 1 MHz for a total of 2 MHz and P-codes spread the signal 10 times that much. Thus we can say the following:
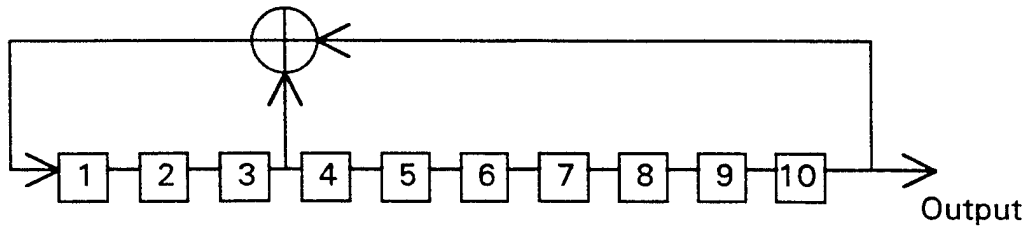
Carrier only $\Rightarrow$ 0 bandwidth
Modulated signal with data $\Rightarrow$ requires minimum bandwidth
Spread spectrum $\Rightarrow$ uses up more bandwidth than needed for data

The spread spectrum property helps GPS to be less susceptible to narrow band jamming, since the jammer energy is concentrated in frequency but the GPS signal energy is spread across such a broad band. Note that this jamming includes both intentional and unintentional kinds.
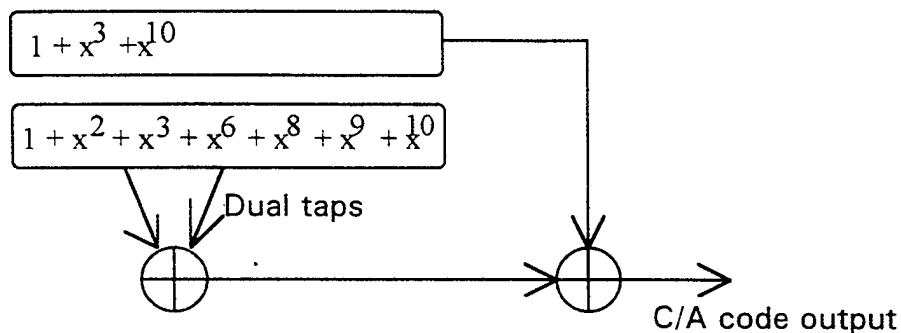
Code Generation

How are the codes generated at the satellite? A GPS satellite generates both the C/A-code and the P-code using devices called tapped feedback shift registers. The left-most cell of a tapped feedback shift register acts as input and accepts a combination of taps from the other cells of the shift register. The output can come from the right-most cell or any point within the register. An infinite number of tapped feedback shift registers are possible. GPS uses a combination of registers yielding high autocorrelations and low cross correlations. In the GPS satellites, two registers are used for C/A-code and 4 registers are used for P-code.

We represent these kinds of registers on paper with polynomials where each exponent represents a tapped cell. For example, C/A-code GPS uses the following tapped feedback shift register for one of the two registers:

$$Polynomial = 1 + x^3 + x^{10}$$

In the tapped feedback shift register above, the input comes from the binary sum of outputs of cells 3 and 10. GPS C/A-code has two registers:

Each GPS satellite has a unique pair of dual taps on the second shift register. Counting up all possible dual taps, we see that 45 possible dual taps exists. Out of these 45 possibilities, GPS uses 32 for the satellites as well as several ground applications.

If we don't limit ourselves to dual taps, but rather use any combination of cells from the second register, then we have a total possibility of

$$2^n - 1 + G_1 + G_2 = 1025$$

codes to choose from. The codes G_1 and G_2 represent the individual shifts registers and the number of cells is $n = 10$. This family of codes we call the gold codes, from which the 32 GPS codes is just a subset.

P-code uses four 10-bit tapped feedback shift registers. Two registers constitute the X1 code and the other two registers make up the X2 code. X1 repeats every 1.5 seconds and X2 has a slightly longer period:

$$X1 \Rightarrow 15,345,000 \quad \textit{chips long}$$
$$X2 \Rightarrow 15,345,037 \quad \textit{chips long}$$

The satellite combines the X1 and X2 codes such that 37 possible delays on X2 produces 37 different P-code segments:
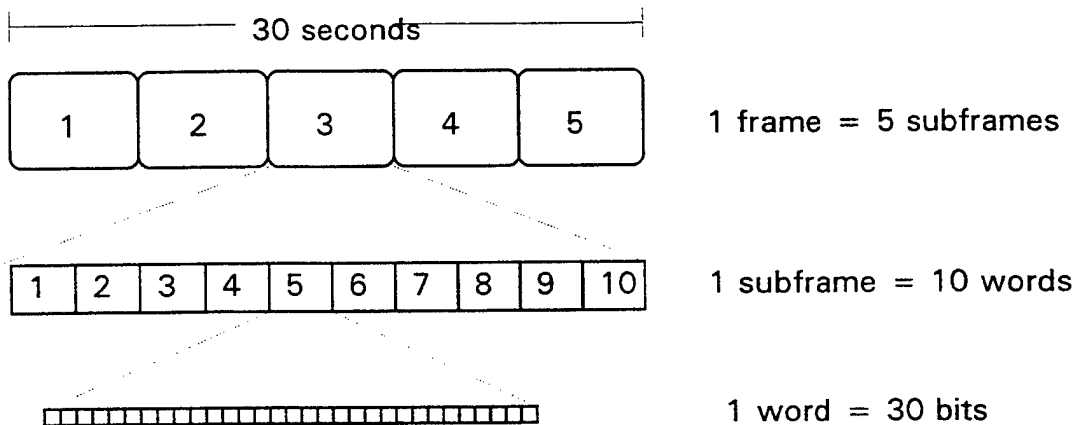
32 for the satellites
5 reserved for ground stations

Data message

The satellite modulates the data onto the L1 and L2 carriers. The following summarizes the important information about the data message:
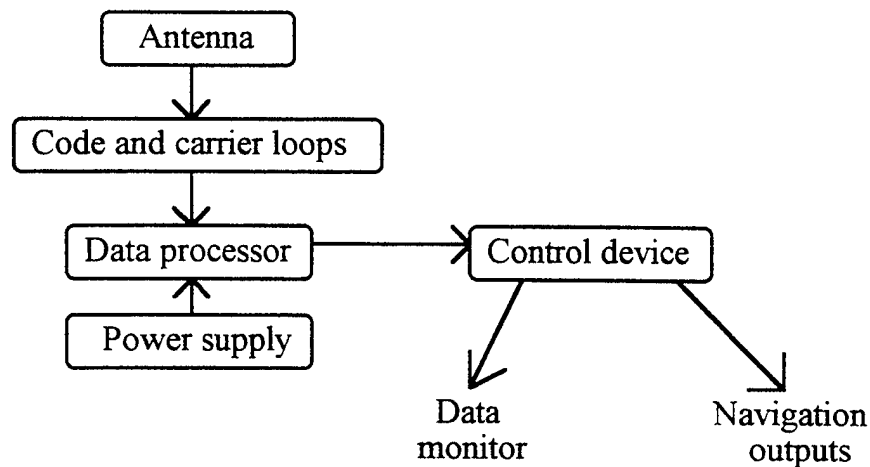
1. 50 bits/second rate (much slower than code or carrier)
2. 30 second frames (1500 bits each frame)
3. Five 6-second subframes per frame
4. HOW transmitted every subframe
5. Clock corrections and ionospheric delay model in subframe #1
6. Ephemeris data in subframes #2 and #3
7. Messages and almanac in subframes #4 and #5

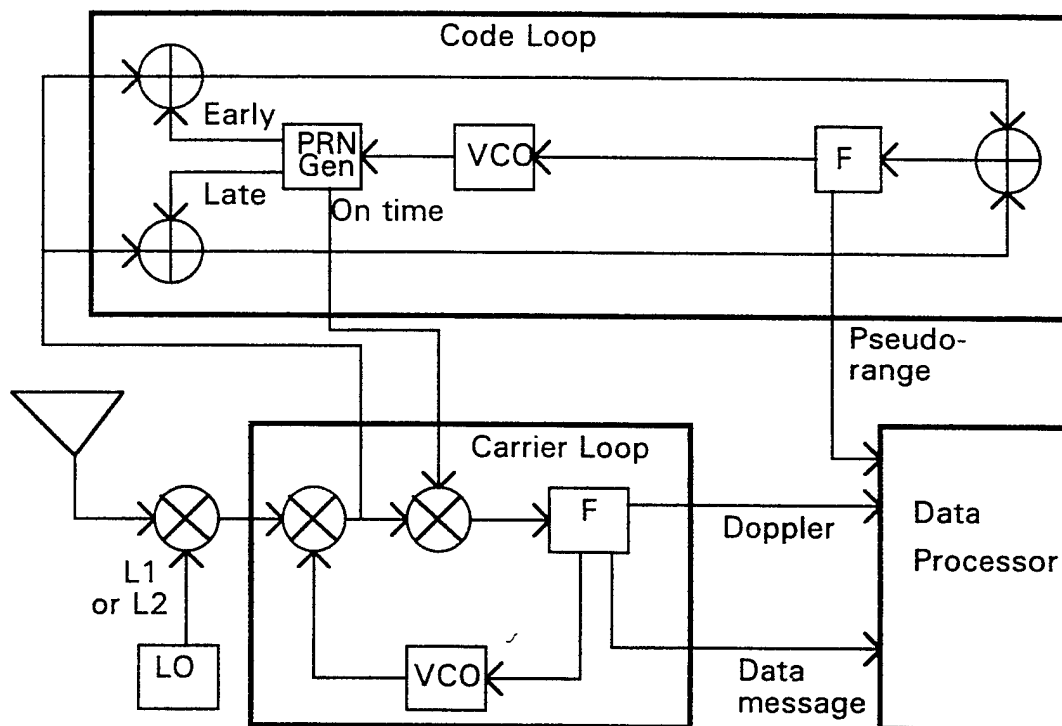The following figure illustrates the format of the data message:

8.2 Inside the Receiver

Now that we've seen how the satellites construct the signal, let's examine how the receiver extracts useful information from the signal. The following figure illustrates the basic building blocks of a typical GPS receiver.

The antenna picks up the GPS signals from all the available GPS satellites. The code loop extracts the code and the carrier loop extracts the data message from each satellite. A typical GPS receiver power source supplies 12 volts D.C. The data processor computes position, velocity, and time given the raw measurements from the code and carrier loops. The control device furnishes interactive communication between the operator and the receiver. A typical receiver allows the operator to choose between the receiver-processed navigation message and raw measurements/ephemeris data (data monitor).

The heart of the receiver is the code tracking loop and the carrier tracking loop. The following figure shows the basic code and carrier loops:

The purpose of the code loop is to provide a measurement of pseudorange for either the C/A-code or P-code. The receiver finds pseudorange by correlating the incoming satellite signal with a delayed internally-generated code. If the internally-generated code and the received code do not align, then the filter F supplies a correction signal to the voltage controlled oscillator (VCO). The VCO controls the receiver generated code. The technique of using the early, late, and on-time internally-generated codes is a popular method in GPS receiver design to maintain code tracking. Once the receiver achieves code lock, it maintains lock. This code loop technique is called a delayed-locked-loop.

The carrier tracking loop demodulates the satellite message by aligning the phase of the channel's local oscillator (LO) with the incoming signal phase. In order to do this, the frequency of the VCO is controlled through a phased-locked-loop. Once the receiver has carrier-lock, the receiver stays locked despite small changes in frequency and phase due to satellite or vehicle movement.

9. *Differential Techniques*

Navigators have long known how differential techniques can significantly reduce navigation errors. For example, standard Omega has errors on the order of several miles. However, by subtracting two measurements with one measurement to a known station, an Omega user can have a position solution with a .25 mile error! The same holds true for GPS. By using differencing we'll be able to eliminate or greatly reduce many of those errors we discussed in Section 6. In other words the common mode errors subtract and therefore cancel out.

How can this be? How do these errors go away or get smaller simply by differencing? To answer that question, consider two zero-mean random variables x and y. These two variables are uncorrelated if and only if

$$E(xy) = E(x)E(y) = 0$$

Now suppose we take their difference

$$z = x - y$$

We measure the strength of z by examining the variance of z:

$$E(z^2) = E\{(x - y)(x - y)\} = E(x^2) - 2E(xy) + E(y^2)$$

If x and y are uncorrelated, then the middle term on the right-hand-side goes away. However, if x and y are correlated, then this term stays:

$$x \text{ and } y \text{ uncorrelated} \Rightarrow E(z^2) = E(x^2) + E(y^2)$$

$$x \text{ and } y \text{ correlated} \Rightarrow E(z^2) = E(x^2) - 2E(xy) + E(y^2)$$

Clearly from these relations, correlated x and y variables cause the variance of their difference to be smaller than for uncorrelated x and y. In fact, in the limiting case of $x = y$ (maximum correlation) we have

$$x \text{ and } y \text{ uncorrelated} \Rightarrow E(x^2) + E(y^2) = 2E(x^2) = 2E(y^2)$$

$$x \text{ and } y \text{ correlated} \Rightarrow E(x^2) - 2E(xy) + E(y^2) = 0$$

Now, consider x and y to be GPS errors on two separate GPS measurements. If these two errors are correlated, then differencing will significantly reduce the strength of the resulting differenced error. If the error is the same for each measurement, then differencing will

eliminate these errors completely. As you will see, this is exactly what happens with GPS differencing techniques!

Actually, we could combine GPS measurements in an infinite number of ways. Here we restrict our attention to differencing between receivers, between satellites, and between times. We could also take individual or single differences, double differences by differencing two single differences, or even triple differences. Below I write out all the possible single/double/triple differences between receivers, satellites, and times:

1. Three types of single differences:

   $\wedge$ Between receivers   (used)
   $\vee$ Between satellites   (used)
   δ Between epochs   (used)

2. Six types of double differences:

   $\wedge\vee$ Between receivers, between satellites   (used)
   $\vee\wedge$ Between satellites, between receivers
   $\wedge\delta$ Between receivers, between epochs   (used)
   $\delta\wedge$ Between epochs, between receivers
   $\vee\delta$ Between satellites, between epochs
   $\delta\vee$ Between epochs, between satellites

3. Six types of triple differences:

   $\wedge\vee\delta$ Between receivers, between satellites, between epochs   (used)
   $\wedge\delta\vee$ Between receivers, between epochs, between satellites
   $\vee\wedge\delta$ Between satellites, between receivers, between epochs
   $\vee\delta\wedge$ Between satellites, between epochs, between receivers
   $\delta\wedge\vee$ Between epochs, between receivers, between satellites
   $\delta\vee\wedge$ Between epochs, between satellites, between receivers

Note that some of these combinations are equivalent. Also, note that only 6 of these differences are commonly used.

Next, I'd like to examine the six commonly used differencing techniques. In this brief analysis you'll see what errors are eliminated or reduced with each technique. Recall from previous sections that the two fundamental measurement equations are

Code-phase measurement equation:

$$(PR_i)_{mod} = \|R_i - R_u\| + c\delta_{ion} + c\delta_{trop} + \delta_{ei} + c\delta_{uc} - c\delta_{sc} + mp_i + sa_i + cdl_i \qquad (9.1)$$

Carrier-phase measurement equation:

$$\left(CP_i\right)_{mod} = \|R_i - R_u\| - c\delta_{ion} + c\delta_{trop} + \delta_{ei} + c\delta_{uc} - c\delta_{sc} + mp_i + sa_i + crl_i + \lambda N \qquad (9.2)$$

These are the basic measurement equations we will modify via differencing in the following subsections. In Equations 9.1 and 9.2 I neglected relativity effects because they are so small and I lumped imaging errors in with multipath. Also *cdl* represents code loop error and *crl* represents carrier loop error.

Between-receivers single differences

Code-phase:

Each between receivers single difference measurement involves two receivers and one satellite at a particular time. Taking the difference between two equations like Equation 9.1 at two different receivers, we get

$$\wedge_i = PR_{i(rec1)} - PR_{i(rec2)}$$
$$\wedge_i = \wedge\|R_i - R_u\| - c\wedge\delta_{uc} + \wedge d_{ion} + \wedge\delta_{trop} + \wedge mp + \wedge cdl$$

Notice that all errors associated with the satellite disappear (ephemeris, selective availability, and satellite clock errors). Also, atmospheric errors greatly reduce if the two receivers are close. Multipath and code loop errors change, but are not necessarily reduced or increased. The first term on the right contains the unknown positions of the two receivers. However, if we nail down one of the receivers on a surveyed spot, then only the position of the other receiver remains as an unknown. This is what we call differential GPS (DGPS). On the other hand, sometimes we don't care about the absolute position of either receiver, but rather their relative positions. In that case, the unknowns become the differences between receiver positions. This is what we call relative GPS (RGPS). In both DGPS and RGPS, navigation solution errors are significantly reduced.

Carrier-phase:

For carrier-phase, we take the difference between two equations like Equation 9.2 and get:
$$\wedge_i = CP_{i(rec1)} - CP_{i(rec2)}$$
$$\wedge_i = \wedge\|R_i - R_u\| - c\wedge\delta_{uc} - \wedge d_{ion} + \wedge\delta_{trop} + \wedge mp + \wedge crl + \lambda\wedge N$$

Again all errors associated with the satellite disappear and atmospheric errors are reduced for small baselines. The biggest change between this equation and the previous code-phase equation is the ambiguity difference. This ambiguity term could be large.

Between-satellites single differences

Code-phase:

Each between satellites single difference measurement involves two satellites and one receiver at a particular time. Taking the difference between two equations like Equation 9.1 for two different satellites, we get

$$\vee = PR_i - PR_j$$
$$\vee = \vee \| R_i - R_u \| + c \vee \delta_{sc} + \vee d_{ion} + \vee \delta_{trop} + \vee mp + \vee cdl + \vee sa$$

Notice that the user clock error disappears! User clock error is a big one, making this method popular. However, atmospheric, multipath, code loop, and sa errors change but are not necessarily reduced.

Carrier-phase:

For carrier-phase, we take the difference between two equations like Equation 9.2 and get:

$$\vee = CP_i - CP_j$$
$$\vee = \vee \| R_i - R_u \| + c \vee \delta_{sc} - \vee d_{ion} + \vee \delta_{trop} + \vee mp + \vee cdl + \vee sa + \lambda \vee N$$

The same comments for code-phase apply here. Again, the biggest change between this equation and the previous code-phase equation is the ambiguity difference, which could be large.

Between-epochs single differences

Code-phase:

Each between epochs single difference measurement involves a single receiver and a single satellite at two different times. Taking the difference between two equations like Equation 9.1 at two different times, we get

$$\delta = PR_{i(time1)} - PR_{i(time2)}$$
$$\delta = \delta \| R_i - R_u \| + c\delta\delta_{sc} - c\delta\delta_{uc} + \delta d_{ion} + \delta\delta_{trop} + \delta mp + \delta cdl + \delta sa$$

This is our pseudorange delta range measurement from Section 4. Notice that all errors could be significantly reduced, depending on the time constant of the errors.

Carrier-phase:

For carrier-phase, we take the difference between two equations like Equation 9.2 and get:

$$\delta = CP_{i(time1)} - CP_{i(time2)}$$

$$\delta = \delta\|R_i - R_u\| + c\delta\delta_{sc} - c\delta\delta_{uc} - \delta d_{ion} + \delta\delta_{trop} + \delta mp + \delta cdl + \delta sa$$

This is our doppler measurement from Section 4. The biggest change between this equation and Equation 9.2 is that the ambiguity term disappears (as long as no cycle slips exist). User velocity is buried in the first term on the right-hand-side. If the errors do not change much over small intervals of time (.001 second typical) then all errors reduce.

Double and triple differences

You can easily write the difference equations for double and triple differences simply by combining the equations for single differences discussed above. The errors that are reduced or eliminated for double and triple differencing are the sum of the reductions and/or eliminations from single differencing. For example, the between receiver-between satellite double difference eliminates satellite errors because of the properties of the between receivers single difference, but also eliminates user clock error because of the properties of between satellite single differences.